



Diachronie en langue de spécialité. Définition d'une méthode linguistique outillée pour repérer l'évolution des connaissances en corpus. Un exemple appliqué au domaine spatial.

Aurélie Picton

► To cite this version:

Aurélie Picton. Diachronie en langue de spécialité. Définition d'une méthode linguistique outillée pour repérer l'évolution des connaissances en corpus. Un exemple appliqué au domaine spatial.. Linguistique. Université Toulouse le Mirail - Toulouse II, 2009. Français. NNT : . tel-00429061

HAL Id: tel-00429061

<https://theses.hal.science/tel-00429061>

Submitted on 30 Oct 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par L'Université Toulouse 2
Discipline ou spécialité : *Sciences du Langage*

Présentée et soutenue par *Aurélie Picton*
Le 20 octobre 2009

Titre

*Diachronie en langue de spécialité.
Définition d'une méthode linguistique outillée pour repérer l'évolution des connaissances
en corpus. Un exemple appliqué au domaine spatial.*

JURY

<i>John Humbley</i>	<i>rapporteur</i>	<i>Professeur, Université Paris 7</i>
<i>Marie-Claude L'Homme</i>	<i>rapporteuse</i>	<i>Professeure, Université de Montréal</i>
<i>Nathalie Aussenac-Gilles</i>	<i>examinatrice</i>	<i>Chargée de Recherche et HDR, CNRS/IRIT</i>
<i>Pascaline Dury</i>	<i>examinatrice</i>	<i>Maître de Conférences, Université Lyon 2</i>
<i>Daniel Galarreta</i>	<i>examineur</i>	<i>Ingénieur cognitif, Cnes</i>
<i>Marie-Paule Péry-Woodley</i>	<i>examinatrice</i>	<i>Professeure, Université Toulouse 2</i>
<i>Anne Condamines</i>	<i>directrice</i>	<i>Directrice de Recherche, CNRS/CLLE-ERSS</i>

École doctorale : **CLESCO**
Unité de recherche : **CLLE-ERSS (UMR5263)**

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 2
Discipline ou spécialité : Sciences du Langage

Présentée et soutenue le 20 octobre 2009 par
Aurélien Picton

Titre

*Diachronie en langue de spécialité.
Définition d'une méthode linguistique outillée pour repérer l'évolution des connaissances en corpus.
Un exemple appliqué au domaine spatial.*

JURY

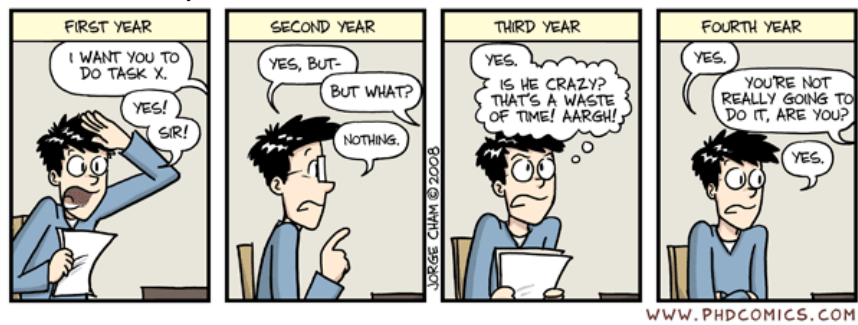
John Humbley	rapporteur	Professeur, Université Paris 7
Marie-Claude L'Homme	rapporteuse	Professeure, Université de Montréal
Nathalie Aussenac-Gilles	examinatrice	Chargée de Recherche et HDR, CNRS/IRIT
Pascaline Dury	examinatrice	Maître de Conférences, Université Lyon 2
Daniel Galarreta	examinateur	Ingénieur cognitif, Cnes
Marie-Paule Péry-Woodley	examinatrice	Professeure, Université Toulouse 2
Anne Condamines	directrice	Directrice de Recherche, CNRS/CLLE-ERSS

École doctorale : CLESCO
Unité de recherche : CLLE-ERSS (UMR5263)

« Welcome. And congratulations. I am delighted that you could make it. Getting here wasn't easy, I know. In fact, I suspect it was a little tougher than you realize. »

(Bill Bryson (2003) *A Short History of Nearly Everything*, Black Swan, Londres)

The Evolution of the "Yes"



("Piled Higher and Deeper" by Jorge Cham, www.phdcomics.com)

Remerciements

On souligne souvent qu'une thèse n'est pas l'œuvre d'une seule personne, et celle-ci ne fait pas exception.

Tout d'abord, je tiens à remercier ma directrice de thèse, Anne Condamines, pour sa patience, son soutien et son aide tout au long de ce travail. Elle a su me guider dans mes réflexions et questionnements (tant théoriques qu'existentiels et qui me paraissaient alors tous insurmontables!) et a su m'encadrer tout en me laissant libre de mes choix. J'espère qu'elle trouvera ici le témoignage de ma reconnaissance.

J'adresse mes plus sincères remerciements à Marie-Claude L'Homme et John Humbley, qui ont eu la gentillesse d'endosser le rôle de rapporteurs. Leurs remarques et commentaires me sont précieux et je suis très honorée de leur implication dans l'évaluation de mon travail.

Un grand merci également à Nathalie Aussenac-Gilles, Pascaline Dury et Marie-Paule Péry-Woodley d'avoir accepté de faire partie de mon jury. Trois personnes qui ont beaucoup compté pour moi ces dernières années : Nathalie pour sa gentillesse, son soutien et ses conseils ; Pascaline, ma « marraine diachronique », pour sa générosité, son entrain et nos échanges ; Marie-Paule pour sa bienveillance, son enthousiasme quotidien et son soutien indéfectible dans tous mes projets TALesques et canadiens.

Rien dans cette recherche n'aurait été possible sans Daniel Galarreta qui m'a ouvert les portes du Cnes et m'a fait partager sa connaissance de ce monde nouveau pour moi. Grâce à lui j'ai eu la chance d'être entourée par une très belle équipe d'experts : George Otrio, Albert Auriol, Jacques Berthon, Josiane Costeraste et Jacques Loesel qui ont accepté de partager leur temps et leurs connaissances pour me permettre de mener à bien cette recherche. Qu'ils en soient tous vivement remerciés.

Je voudrais également dire ici ma reconnaissance d'avoir été soutenue tout au long de ce travail par mon équipe d'accueil, CLLE-ERSS, qui prend remarquablement soin de sa « relève ». J'ai beaucoup appris dans cette équipe et y ai été vraiment très bien entourée. Un grand merci aux Erssiens, dont bon nombre ont été mes enseignants, et qui ont tous eu à un moment donné un mot gentil, un conseil ou une référence à me glisser à l'oreille. Je voudrais remercier tout particulièrement : Michel Aurnague, Andrée Borillo, Didier Bourigault (et Saint-Exupéry), Myriam Bras, Anne Condamines, Jacques Durand, Nabil Hathout, Amélie Josselin-Leray, Anne Le Draoulec, Fabio Montermini, Marie-Paule Péry-Woodley, Ludovic Tanguy et Jean-Michel Tarrier. Deux mentions spéciales : la première pour Cécile Fabre, ma première enseignante de linguistique et ma « bonne fée » ; la seconde pour Josette Rebeyrolle qui a été une relectrice, un soutien et un guide sans égal ces derniers mois. Je lui dois beaucoup. Une belle pensée également pour Laurence Lamy et Nathalie Moulic qui m'ont beaucoup facilité la vie et avec qui j'ai partagé de très bons moments. Un grand merci

aussi à Corinne Prunier pour sa disponibilité, à Franck Sajous pour sa gentillesse et ses connaissances informatiques et à Bruno Chenu et Corine Ratier pour leur soutien technique.

Un autre laboratoire qui a beaucoup compté pour moi est l'Observatoire de linguistique Sens-Texte de l'Université de Montréal qui m'a accueillie à bras ouverts à plusieurs reprises pour mes recherches. Merci entre autres à Marie-Claude L'Homme, Patrick Drouin (pour sa relecture de dernière minute aussi...) et à leurs étudiants, parmi eux Iveth Carreño, Anne-Laure Jousse et Annaïch Le Serrec.

Enfin, une énorme pensée aux doctorants du laboratoire CLLE-ERSS avec qui j'ai partagé bien plus que des bureaux. J'espère que l'avenir nous permettra de nous retrouver souvent pour partager fous-rires, thé, repas-de-11h30-au-resto, soirées, concours de Noël, chansons et apéros mais aussi des réflexions et des recherches (il ne faudrait pas que ceux qui lisent ces remerciements pensent que l'on ne fait rien de sérieux au 6^{ème}!). Cette ambiance et ces excellents moments je les dois en particulier (par ordre d'apparition) à : Marianne Vergez-Couret (et son si joli rire), Christel Le Bellec, Frédérique Saez, Christelle Pêcher (quasi ERSS), Buddy Dirat (t'inquiète), Nathalie Dehaut (et Ray), Aurélie Guerrero, Stéphanie Lopez, Clémentine Adam et Lionel Fontan. 1000 mercis également à Christine Fèvre-Pernet et Pascale Vergely pour leurs relectures attentives (et vivantes !) et leurs encouragements, ainsi qu'à Mai Ho-Dac, Valérie Braun, Anh Ngoc Vo, Marion Laignelet, Vladimir Moscovici, Christophe Pimm, Stéphanie Volteau et John Giraldo.

Pour finir, je tiens à remercier de tout cœur mon entourage qui a tenu bon toutes ces années et a souvent cherché à comprendre ce qui pouvait bien m'occuper et qui aurait un rapport avec la langue et le spatial... Merci en particulier à ma sœur, à Deok-Hee et Olivier Dufor, Ganaël Jatteau, Guillaume Lamontagne et Franck Pontiers d'avoir été là, de m'avoir supportée, relue, sortie de mon ordinateur, nourrie, chouchoutée, soutenue et tant de choses impossibles à lister. Mais parmi mes proches, je voudrais dédier ce travail à mes parents qui ont toujours cru en moi et qui m'ont toujours accompagnée quelles que soient mes folies, ainsi qu'à Rafaëlle Delepaut, meilleure amie, globe-trotteuse, relectrice de haut rang et championne du monde de conversation Skype par équipe : Ma toi, je n'y serais jamais arrivée sans toi.

Je tiens également à remercier les organismes suivants, qui m'ont accordé le soutien financier nécessaire à la réalisation de cette recherche : le Centre National d'Études Spatiales, le Centre National de la Recherche Scientifique, le Conseil Scientifique de l'Université Toulouse 2 et le Centre de Coopération Inter-universitaire Franco-Québécois.

Table des matières

REMERCIEMENTS	VII
TABLE DES MATIÈRES	IX
LISTE DES TABLEAUX ET DES FIGURES	XV
INTRODUCTION GÉNÉRALE	1
 PREMIÈRE PARTIE – POUR UNE APPROCHE DIACHRONIQUE EN LANGUE DE SPÉCIALITÉ : ÉLÉMENTS THÉORIQUES ET MÉTHODOLOGIQUES.....	 7
CHAPITRE 1 DIACHRONIE ET ÉVOLUTION DANS LA LANGUE.....	9
1.1 PERSPECTIVE DIACHRONIQUE ET LANGUE GÉNÉRALE	9
1.1.1 Une « tradition » linguistique	10
1.1.2 Diachronie en langue générale : Trois caractéristiques	11
1.1.2.1 Diversité des niveaux de description	11
1.1.2.2 Facteurs internes et externes du changement sémantique	12
1.1.2.3 Place des questions méthodologiques	14
1.2 PERSPECTIVE DIACHRONIQUE ET LANGUE DE SPÉCIALITÉ	17
1.2.1 Fondements de la terminologie et diachronie	18
1.2.2 Descriptions diachroniques en langue de spécialité : bilan en demi-teinte	19
1.2.3 Émergence de nouvelles perspectives diachroniques.....	23
1.2.3.1 Nouvelles théories de la terminologie	23
1.2.3.2 Néologie, implantation et terminochronie : Trois approches diachroniques	28
1.2.3.3 Terminologie computationnelle et applications.....	35
1.3 SYNTHÈSE	38
CHAPITRE 2 POUR UN PARALLÈLE ENTRE ÉVOLUTION DE LA LANGUE ET ÉVOLUTION DES CONNAISSANCES	41
2.1 INTERNE/EXTERNE : UNE DISTINCTION THÉORIQUE FONDAMENTALE EN TERMINOLOGIE	42
2.1.1 Définition du terme	42
2.1.1.1 Du « concept et terme-étiquette »... ..	42
2.1.1.2 ... au « terme-signe linguistique et concept »	44
2.1.1.3 Appartenance à un domaine.....	45
2.1.2 Nouveaux rapports entre textes et connaissances	48
2.1.2.1 Terminologie textuelle	48
2.1.2.2 Dimension externe : définir les « connaissances »	51
2.1.2.3 Dimension interne : la langue et les textes	54
2.2 POINTS DE VUE HISTORIQUE ET DIACHRONIQUE.....	55
2.2.1 Histoire interne et histoire externe	55
2.2.2 Hypothèse d'un parallèle entre langue et connaissances pour la diachronie	59

CHAPITRE 3 MÉTHODOLOGIE POUR UNE ANALYSE DE L'ÉVOLUTION EN CORPUS61

3.1	PRINCIPES MÉTHODOLOGIQUES GÉNÉRAUX POUR L'ANALYSE DE L'ÉVOLUTION.....	62
3.1.1	Démarche comparative : Périodisation et organisation temporelle des corpus	62
3.1.1.1	Synchronie dynamique et diachronie	62
3.1.1.2	Organisation temporelle des corpus : Choix des périodes et intervalles	63
3.1.1.3	Trois types de corpus : Corpus synchroniques, corpus de suivi et corpus comparables.....	67
3.1.2	Démarche outillée : Linguistique « à l'instrument » pour observer le changement	70
3.1.2.1	Linguistique outillée : de nouveaux moyens et de nouveaux besoins	70
3.1.2.2	Outillage et indices.....	73
3.2	CHOIX MÉTHODOLOGIQUES POUR NOTRE ÉTUDE	77
3.2.1	Corpus pour le domaine spatial	77
3.2.1.1	Choix des sources	77
3.2.1.2	Domaine d'activité et domaine de connaissance.....	79
3.2.1.3	Contraintes de constitution des corpus diachroniques	82
3.2.2	Description des corpus sélectionnés.....	84
3.2.2.1	Corpus TTVS.....	84
3.2.2.2	Corpus DORIS.....	88
3.2.3	Choix des outils et traitement des corpus	91
3.2.3.1	Choix des outils.....	91
3.2.3.2	Extraction et préparation des données	97

DEUXIÈME PARTIE – EXPLORATION OUTILLÉE DE QUATRE INDICES LINGUISTIQUES 103

CHAPITRE 4 EMPREINTES DE FRÉQUENCE105

4.1	EMPREINTES DE FRÉQUENCE ET CORPUS DIACHRONIQUES.....	106
4.1.1	« Empreintes de fréquence » : définition.....	106
4.1.2	Néologie et obsolescence.....	107
4.1.3	Formes chrono-homogènes.....	109
4.2	MÉTHODOLOGIE POUR LA MISE AU JOUR D'EMPREINTES.....	110
4.2.1	Sélection du seuil de fréquence	110
4.2.2	Définition d'empreintes de fréquence : mise en œuvre du χ^2	111
4.3	OBSERVATION DES EMPREINTES DE FRÉQUENCE EN CORPUS.....	115
4.3.1	Apparition et disparition de termes.....	115
4.3.1.1	Corpus TTVS.....	115
4.3.1.2	Corpus DORIS.....	124
4.3.1.3	Synthèse.....	130
4.3.2	Termes chrono-homogènes.....	132
4.3.2.1	Repérage de thématiques	132
4.3.2.2	Interprétations des experts.....	134
4.3.2.3	Synthèse.....	142
4.4	BILAN.....	144

CHAPITRE 5 CONTEXTES RICHES EN CONNAISSANCES ÉVOLUTIVES147

5.1	CONTEXTES RICHES EN CONNAISSANCES ÉVOLUTIVES ET MARQUEURS D'ÉVOLUTION	148
5.1.1	Définition.....	148
5.1.2	Exemples d'extraction d'informations évolutives dans les textes	149
5.2	EXTRACTION DES CONTEXTES D'ÉVOLUTION À L'AIDE DE MARQUEURS.....	158

5.2.1	<i>Marqueurs définis pour cette recherche</i>	158
5.2.2	<i>Extraction de contextes riches en connaissances évolutives</i>	168
5.3	OBSERVATIONS EN CORPUS : PROJECTION DES TERMES EN CONTEXTES	169
5.3.1	<i>Évaluation de la pertinence des contextes et des marqueurs</i>	169
5.3.2	<i>Observations de contextes</i>	177
5.3.2.1	Point de vue synchronique dynamique.....	178
5.3.2.2	Point de vue diachronique	180
5.3.3	<i>Observations de termes récurrents en contextes</i>	182
5.4	BILAN.....	185
CHAPITRE 6 VARIANTES TERMINOLOGIQUES		189
6.1	VARIANTES TERMINOLOGIQUES ET DIACHRONIE.....	190
6.1.1	<i>Définition et richesse des perspectives diachroniques</i>	190
6.1.2	<i>Entre langue et discours</i>	193
6.2	EXTRACTION ET SÉLECTION DES VARIANTES TERMINOLOGIQUES	196
6.2.1	<i>Typologie définie pour cette étude</i>	196
6.2.2	<i>Sélection des variantes</i>	199
6.3	OBSERVATION EN CORPUS	206
6.3.1	<i>Évaluation de la pertinence des variantes terminologiques en diachronie</i>	206
6.3.2	<i>Apparition simultanée de variantes</i>	207
6.3.3	<i>Apparition/disparition de certaines variantes</i>	209
6.3.4	<i>Variants lexicales et variantes discursives : recours aux contextes pour l'analyse</i>	214
6.4	BILAN.....	219
CHAPITRE 7 DÉPENDANCES SYNTAXIQUES.....		221
7.1	DÉPENDANCES SYNTAXIQUES ET ANALYSE DIACHRONIQUE	221
7.1.1	<i>Dépendances et cooccurrences : définitions</i>	221
7.1.2	<i>Exploration des dépendances en diachronie</i>	224
7.2	EXTRACTION DES DÉPENDANCES SYNTAXIQUES EN CORPUS	231
7.2.1	<i>Analyse syntaxique automatique de l'outil Syntex</i>	231
7.2.2	<i>Sélection des dépendances</i>	232
7.3	OBSERVATIONS EN CORPUS.....	236
7.3.1	<i>Évaluation de la pertinence globale des dépendances syntaxiques en diachronie</i>	236
7.3.2	<i>Nouveauté et obsolescence : productivité morphologique et lexicale</i>	238
7.3.3	<i>Apparition/disparition de thématiques</i>	240
7.3.4	<i>Changement de point de vue</i>	243
7.4	BILAN.....	249
CHAPITRE 8 TYPOLOGIE DE L'ÉVOLUTION EN DIACHRONIE COURTE.....		253
8.1	TYPOLOGIE DE L'ÉVOLUTION EN DIACHRONIE COURTE.....	254
8.2	RETOUR SUR LES DIMENSIONS INTERNE ET EXTERNE	261
8.2.1	<i>Dénomination, concept, instance</i>	261
8.2.2	<i>Différents niveaux d'évolution</i>	262
8.2.3	<i>Évolution des techniques et évolution des sciences</i>	264
8.2.4	<i>Évolution et continuité</i>	267
8.2.4.1	Entre discours et langue	267
8.2.4.2	Du même à l'autre	268
8.3	CONSÉQUENCES MÉTHODOLOGIQUES.....	270

TROISIÈME PARTIE – CONSTRUIRE L’INTERPRÉTATION : FAISCEAUX D’INDICES ET PRINCIPE DE CO-CONSTRUCTION	273
CHAPITRE 9 INTERPRÉTATION DE L’ÉVOLUTION VIA DES « FAISCEAUX » D’INDICES	275
9.1 COMBINER LES INDICES POUR GUIDER L’INTERPRÉTATION.....	275
9.1.1 <i>L’exemple de MOS/CMOS/BiCMOS</i>	276
9.1.2 <i>L’exemple de la thématique du refroidissement</i>	279
9.1.3 <i>L’exemple de l’influence du domaine informatique dans DORIS</i>	281
9.2 ENTRETIENS AUPRÈS D’EXPERTS	285
9.2.1 <i>Entretiens préliminaires</i>	285
9.2.1.1 Tâche proposée	286
9.2.1.2 Résultats.....	288
9.2.2 <i>Seconds entretiens : éléments d’évaluation</i>	291
9.2.2.1 Tâche proposée	291
9.2.2.2 Résultats.....	292
9.2.2.3 Bilan et limites de l’évaluation.....	295
9.2.2.4 Évaluation sur le corpus DORIS	297
9.3 BILAN.....	297
CHAPITRE 10 POUR UNE CO-CONSTRUCTION DE L’INTERPRÉTATION	299
10.1 PRINCIPE DE CO-CONSTRUCTION.....	300
10.1.1 <i>Construire l’interprétation</i>	300
10.1.2 <i>Consensus interprétatif</i>	301
10.2 ACTEURS DE LA CO-CONSTRUCTION	303
10.2.1 <i>Analyste</i>	303
10.2.1.1 Définition de l’analyste.....	303
10.2.1.2 Atouts : l’analyse de corpus de textes	303
10.2.1.3 Limites de l’analyste.....	304
10.2.2 <i>Expert du domaine</i>	305
10.2.2.1 Rôle de l’expert	305
10.2.2.2 Qualités et atouts d’un expert pour la diachronie.....	306
10.2.2.3 Difficultés du « rétro-diagnostic »	307
10.2.3 <i>Nature des échanges</i>	313
10.3 PRISE EN COMPTE DE L’APPLICATION	314
10.3.1 <i>Application comme guide de l’analyse</i>	314
10.3.2 <i>L’exemple de la mise à jour de ressources termino-ontologiques</i>	316
CONCLUSION GÉNÉRALE	321
RÉFÉRENCES BIBLIOGRAPHIQUES	329

INDEX THÉMATIQUE	365
INDEX DES AUTEURS	369
INDEX DES TERMES/CONCEPTS TRAITÉS	373
 ANNEXES	 377
 ANNEXE A – EXTRAITS DES CORPUS	 379
CORPUS TTVS	379
CORPUS DORIS	381
ANNEXE B – PRÉCISION DES MARQUEURS D'ÉVOLUTION (TTVS)	383
ANNEXE C – TYPOLOGIE DE L'ÉVOLUTION EN DIACHRONIE COURTE	386

Liste des tableaux et des figures

Tableau 3.1 – Nombre d'occurrences – Corpus et sous-corpus TTVS	87
Tableau 3.2 – Nombre d'occurrences – Corpus et sous-corpus DORIS.....	90
Tableau 3.3 – Description des corpus en fonction des contraintes de constitution	91
Tableau 3.4 – Réduction du nombre de CT après filtrage – corpus TTVS	101
Tableau 3.5 – Réduction du nombre de CT après filtrage – corpus DORIS.....	101
Tableau 4.1 – Table de contingence – Calcul du χ^2 du terme <i>polissage</i> (TTVS)	112
Tableau 4.2 – Extrait du tableau de distribution du χ^2 (Muller, 1968 : 179).....	113
Tableau 4.3 – Écarts obtenus – exemple de <i>polissage</i> (TTVS).....	114
Tableau 4.4 – Exemples pour chaque empreinte de fréquence.....	115
Tableau 4.5 – Exemples de termes apparus (TTVS)	115
Tableau 4.6 – Exemples de termes disparus (TTVS).....	116
Tableau 4.7 – 7 « Candidats néologismes » significatifs ($\chi^2 \geq 9,210$ et freq ≥ 30) –TTVS	117
Tableau 4.8 – 5 « Candidats obsolètes » (5 termes les plus fréquents disparus) –TTVS.....	117
Tableau 4.9 – Exemples de termes apparus (DORIS).....	124
Tableau 4.10 – Exemples de termes disparus (DORIS)	125
Tableau 4.11 – 18 « Candidats néologismes » significatifs ($\chi^2 \geq 6,635$ et freq ≥ 20) – DORIS	126
Tableau 4.12 – 9 « Candidats obsolètes » significatifs ($\chi^2 \geq 6,635$ et freq ≥ 20) – DORIS.....	126
Tableau 4.13 – 6 empreintes de fréquence (TTVS).....	133
Tableau 4.14 – 2 empreintes de fréquence (DORIS)	133
Tableau 4.15 – Répartition des termes chrono-homogènes en thématiques (TTVS)	135
Tableau 4.17 - Répartition des termes chrono-homogènes en thématiques (DORIS).....	140
Tableau 5.1 – Exemples de marqueurs (Ibekwe-Sanjuan, 2005 : 266)	151
Tableau 5.2 – Exemples de marqueurs (Sándor, 2006).....	153
Tableau 5.3 – Exemples de marqueurs (Laignelet, 2009)	155
Tableau 5.4 – Pertinence des contextes d'évolution.....	170
Tableau 5.5 – Répartition des contextes en fonction du nombre de marqueurs	172
Tableau 5.6 – Nombre et proportion de contextes pertinents en fonction du nombre de marqueurs	172
Tableau 5.7 – Pertinence des contextes d'évolution.....	173
Tableau 5.8 – Précision des trois groupes principaux de marqueurs (TTVS).....	174
Tableau 5.9 – Précision de chaque marqueur dans l'ensemble du corpus (TTVS)	176
Tableau 5.10 – Précision des marqueurs en fonction de leur classe dans chaque sous-corpus	177
Tableau 6.1 – Exemple de groupe de variantes en concurrence (DORIS)	201
Tableau 6.2 – Exemple de groupe de variantes dont au moins une apparaît/disparaît (TTVS)	201
Tableau 6.3 – Exemple de calcul du coefficient d'implantation (TTVS).....	202
Tableau 6.4 – Sélection des variantes (TTVS).....	203
Tableau 6.5 – Sélection des variantes (DORIS).....	204
Tableau 6.6 – Répartition des variantes sélectionnées en diachronie (TTVS et DORIS)	205

Tableau 6.7 – Pertinence des variantes sélectionnées en diachronie (TTVS)	207
Tableau 6.8 – Pertinence des variantes sélectionnées en diachronie (DORIS).....	207
Tableau 6.9 – Coefficient d'implantation des variantes de <i>synthèse d'ouverture optique</i> (TTVS)	210
Tableau 6.10 – Coefficient d'implantation des variantes de <i>liaisons entre satellites</i> (TTVS) ...	211
Tableau 6.11 – Coefficient d'implantation des variantes liées à l'infrarouge (TTVS)	211
Tableau 6.12 – Coefficient d'implantation des variantes de <i>panne</i> (DORIS)	213
Tableau 7.1 – Principales dépendances syntaxiques observées.....	231
Tableau 7.2 – Exemples de descendants apparus simultanément – <i>polissage</i> (TTVS).....	233
Tableau 7.3 – Exemples de quelques descendants apparus – <i>laser</i> (TTVS)	233
Tableau 7.4 – Sélection des termes têtes qui présentent des descendants (TTVS et DORIS)	234
Tableau 7.5 – Répartition des descendants en fonction des empreintes de fréquence (TTVS et DORIS)	235
Tableau 7.6 – Répartition des termes têtes en fonction des empreintes de fréquence de leurs descendants (TTVS et DORIS).....	235
Tableau 7.7 – Pertinence des séries paradigmatiques (TTVS).....	237
Tableau 7.8 – Pertinence des séries paradigmatiques (DORIS)	237
Tableau 7.9 – Exemples de productivité morphologique et lexicale (apparition de concepts – TTVS).....	238
Tableau 7.10 – Exemple de productivité morphologique et lexicale (disparition d'une fonctionnalité – DORIS).....	239
Tableau 7.11 – Exemple de productivité morphologique et lexicale (disparition d'instances – <i>touche</i>) (DORIS).....	239
Tableau 7.12 – Exemple de productivité morphologique et lexicale (apparition de fonctionnalité – <i>visibilité satellite</i>) (DORIS)	240
Tableau 7.13 – Apparition d'une série paradigmatique – thématique des miroirs (TTVS) ..	241
Tableau 7.14 – Apparition d'une série paradigmatique – thématique des sondeurs (TTVS)	242
Tableau 7.15 – Exemple de changement de point de vue – le cas d' <i>opérateur</i> (DORIS).....	245
Tableau 7.16 – Exemple de changement de point de vue – le cas de <i>parasite</i> (TTVS).....	246
Tableau 7.17 – Changement de point de vue – <i>lumière parasite</i> (TTVS)	248
Tableau 9.1 – Productivité morphologique et lexicale – MOS/CMOS/BiCMOS.....	276
Tableau 9.2 – Variantes graphiques bi-CMOS/BiCMOS.....	277
Tableau 9.3 – Apparition des nouveaux descendants du terme <i>refroidir</i> (TTVS).....	280
Tableau 9.4 – Extrait des dépendances syntaxiques d' <i>interface</i> (DORIS).....	282
Tableau 9.5 – Extrait des dépendances syntaxiques d' <i>opérateur</i> (DORIS)	283
Tableau 9.6 – Jugement des experts sur la seconde liste.....	292
Tableau 9.7 – Nombre et proportion de termes concernés par les indices.....	293
Tableau 9.8 – Nombre et proportion de termes pertinents concernés par les indices.....	294
Tableau 9.9 – Nombre et proportion de termes concernés par les variantes vs. les contextes	295
Tableau 9.10 – Nombre et proportion de termes pertinents concernés par les variantes vs. les contextes	295

Figure 3.1 – Extrait de la liste Syntex de candidats-termes – Interface TermOnto	94
Figure 3.2 – Extrait de l'analyse syntaxique des dépendances (TermOnto)	95
Figure 3.3 – Interface de visualisation (TerminoWeb)	96
Figure 4.1 – Calcul du χ^2 (Howell, 2008 : 143)	112
Figure 6.1 – Coefficient d'implantation (Quirion, 2003a ; 2003b)	202
Figure 7.1 – Schéma d'un réseau de dépendances Syntex (extrait DORIS).....	232
Figure 9.1 – Extrait de la liste présentée (premiers entretiens)	288
Figure 9.2 – Extrait de la liste présentée (seconds entretiens).....	291
Figure 10.1 – Schéma de co-construction de l'interprétation	302

Introduction générale

Dans cette thèse, nous menons une réflexion sur la place de la dimension diachronique dans les langues de spécialité, à travers la définition d'une méthode linguistique pour repérer l'évolution des connaissances en corpus.

En terminologie et en langue de spécialité, la dimension diachronique est très largement ignorée, ce que regrettent de nombreux chercheurs (tels que Dury, 1997 ; Guespin, 1995 ; Humbley, 1994 ; Møller, 1998 ; Sager, 1990 et Van Campenhoudt, 1998). Cette situation s'explique par plusieurs raisons, dont les deux principales sont d'ordre théorique et technique (Dury & Picton, à paraître). D'un point de vue théorique, la terminologie a en effet longtemps été dominée par la perspective classique wüsterienne, dont l'objectif premier de normalisation exclut la variation temporelle, autrement dit la dimension diachronique ; cela explique que, par conséquent, d'un point de vue technique, très peu de ressources et d'outils existent aujourd'hui pour aborder cette perspective. Néanmoins, depuis les années 1990, la terminologie voit ses fondements théoriques rediscutés et de nouvelles propositions apparaissent qui favorisent la description du fonctionnement du terme, dans toute sa variation. Cette remise en cause est accompagnée et soutenue par les progrès de la linguistique de corpus et de la linguistique outillée, qui seuls permettent d'explorer de nouvelles voies de recherches jusqu'alors écartées, telles que la diachronie.

Mais bien que la prise de conscience de l'importance de la diachronie soit aujourd'hui réelle, le manque d'études sur l'évolution des connaissances et des langues de spécialité est patent. Or, les connaissances scientifiques et techniques évoluent rapidement et avec elles les dénominations qui les désignent. Ces changements, ni le terminologue ni les experts de domaine ne peuvent les ignorer. Ainsi, comme le note Rey (1992 : 58), l'expert a nécessairement besoin d'être tenu au courant de l'évolution récente de son domaine, ce qui implique « de maîtriser les notions et les termes nouveaux, ainsi que les modifications dans l'affectation des termes à des notions en évolution plus rapide qu'eux ». Il est donc temps de rattraper le retard dans les descriptions et de proposer des méthodes qui permettent une

analyse fine et fiable des phénomènes diachroniques en jeu. Notre recherche se situe précisément dans cette optique.

Le besoin de description est d'ailleurs soutenu par nombre de perspectives appliquées. C'est ce que montre notamment Dury (1997 ; 2008) qui insiste sur l'importance pour les traducteurs d'avoir accès à des informations sur l'évolution des concepts afin d'éviter les pièges de traduction que le changement entraîne, ou encore Celotti & Musacchio (2004) qui démontrent l'intérêt de décrire l'évolution des termes pour la didactique des langues de spécialité. Aujourd'hui, l'une des demandes les plus fortes provient certainement des applications liées à la gestion des connaissances (Aussenac-Gilles *et al.*, 2007 ; 2007b) et parmi elles, la mise au jour de ressources termino-ontologiques. Ces ressources (thésaurus, glossaires, bases de données terminologiques ou encore ontologies) formalisent en effet les connaissances d'un domaine à une période donnée. Or, les connaissances évoluant, ces ressources deviennent obsolètes. Leur mise à jour constitue donc un enjeu majeur pour lequel il faut mettre en place une étude diachronique pour repérer le changement.

Notre démarche s'ancre dans ce type de demandes appliquées, et plus précisément dans une demande formulée par le Centre National d'Études Spatiales¹ (Cnes)². Dans le domaine spatial, la question de l'évolution est cruciale dans le cadre de la mise en place de projets spatiaux dits « de longue durée » (c'est-à-dire une période d'une dizaine à une trentaine d'années). Ces projets concernent par exemple l'envoi dans l'espace de sondes qui n'atteignent leur destination que plusieurs années après leur lancement, telle que la sonde Rosetta lancée en mars 2004 et qui n'atteindra son objectif (la comète Churyumov-Gerasimenko) qu'en 2014³. De la même manière, plusieurs générations d'un satellite ou d'un instrument peuvent être développées et nécessitent un suivi de la part des experts, comme dans le cas de la filière Spot dont les satellites se succèdent sur 5 générations depuis 1986. Dans ces contextes, entre le début et la fin d'un projet spatial ou entre la première et la dernière génération d'un instrument, les connaissances impliquées évoluent inévitablement

¹ Notre recherche bénéficie d'un financement conjoint du Centre National d'Études Spatiales (Cnes) et du Centre National de la Recherche Scientifique (CNRS) (Bourse de Docteur Ingénieur).

² <http://www.cnes.fr>.

³ <http://smc.cnes.fr/ROSETTA/Fr/> (consulté le 3 avril 2009).

et le plus souvent à l'insu des experts. De fait, ces conditions entraînent de nombreuses difficultés dans leur pratique : mauvaise communication entre ingénieurs « en poste » et ingénieurs « juniors » qui arrivent en cours de projet, oubli du contexte de connaissances dans lequel le projet a été initié, plus généralement, modification non consciente du sens et/ou de la forme des termes. Face à ce constat, le Cnes a clairement identifié ces questions et mis en avant la nécessité d'apprendre à repérer l'évolution (Galarreta, 2001 ; Rothenburger & Galarreta, 2004).

Notre objectif consiste à mettre en place une méthode linguistique de repérage de l'évolution des connaissances en corpus qui réponde à ces besoins appliqués. Néanmoins, cet ancrage appliqué ne peut pas être isolé de questionnements théoriques et descriptifs sur la prise en compte de la dimension diachronique en langue de spécialité, questionnements sur lesquels nous souhaitons proposer une réflexion originale.

Afin d'atteindre cet objectif, nous prenons appui sur l'expérience de la terminologie textuelle pour l'analyse terminologique. Nous reprenons à notre compte les principes selon lesquels, d'une part, les connaissances partagées par les experts sont accessibles dans les textes spécialisés et d'autre part, les applications liées à la terminologie sont étroitement liées aux textes (Bourigault & Slodzian, 1999 : 30). Cette position nous permet de poser un parallèle entre langue et connaissances pour observer l'évolution : puisque c'est dans les textes que sont exprimées les connaissances, repérer des évolutions linguistiques dans les textes constitue la clé pour repérer des évolutions de connaissances dans le domaine.

Pour valider cette hypothèse, trois étapes s'imposent. Tout d'abord, il s'agit d'identifier et de définir des indices linguistiques repérables en corpus que l'on peut associer de manière régulière à des évolutions de connaissances. La mise au jour de ces indices repose sur une exploration linguistique outillée de corpus de textes.

Deuxièmement, il est crucial de montrer en quoi chaque indice contribue au repérage d'évolutions potentielles et comment on peut les mettre en œuvre. Plus spécifiquement, il s'agit de préciser comment articuler le recours aux outils, les indices eux-mêmes ainsi que l'interprétation conjointe de l'analyste et de l'expert pour associer évolution dans la langue et évolution des connaissances.

Enfin, afin de juger de la pertinence de la méthodologie proposée, nous définissons une typologie des phénomènes d'évolution des connaissances ainsi repérés grâce à notre approche.

C'est de l'articulation de ces trois étapes que dépend, selon nous, la qualité de la réflexion et des réponses que l'on pourra apporter à la question de la place de la diachronie en langue de spécialité. Pour le dire autrement, il s'agit d'articuler un triple regard : un regard théorique, un regard descriptif et un regard appliqué.

D'un point de vue appliqué, notre contribution se situe dans la mise au jour d'une méthode linguistique, reproductible et fiable pour repérer l'évolution des connaissances en corpus. D'un point de vue descriptif, notre apport se situe dans l'étude de phénomènes d'évolution en jeu en diachronie, développée dans un contexte original dit « en diachronie courte », c'est-à-dire la description de l'évolution sur de très courts intervalles temporels. Enfin, d'un point de vue théorique, notre contribution repose sur la manière d'ancrer l'analyse diachronique en terminologie textuelle et sur l'impact qu'un tel point de vue implique sur le lien entre textes et connaissances et sur la spécificité de l'analyse diachronique dans les langues de spécialité.

Pour développer ces différents aspects et exposer notre réflexion, nous organisons notre manuscrit en trois parties. Dans la première partie, nous posons les éléments théoriques et méthodologiques fondamentaux pour aborder une étude diachronique. Le chapitre 1 situe notre recherche à travers un panorama des différentes approches traitant de l'évolution et sur lesquelles construire une étude diachronique en langue de spécialité. Le chapitre 2 présente notre ancrage en terminologie textuelle et les fondements du parallèle entre langue et connaissances, sur lequel nous nous appuyons. Nous y montrons également la manière dont ce parallèle permet de poser les jalons d'une exploration linguistique en corpus inhérente à notre approche. Dans le chapitre 3, nous détaillons les principaux choix méthodologiques auxquels est confronté l'analyste devant une telle entreprise. Nous nous concentrons en particulier sur la question de la constitution de corpus diachroniques et de la mise en place d'une approche linguistique outillée pour l'exploration. Nous justifions ces différents points méthodologiques à travers l'exemple de la constitution des deux corpus construits pour notre analyse.

La deuxième partie est consacrée à la définition et à l'exploration en corpus d'indices linguistiques pour le repérage de l'évolution. Quatre indices sont repérés : les empreintes de fréquence, les contextes riches en connaissances évolutives, les variantes terminologiques et les dépendances syntaxiques. Les chapitres 4 à 7 détaillent chaque indice et la manière dont ils doivent être mis en œuvre en diachronie. Nous évaluons également leur pertinence à travers la description fine de l'évolution des connaissances qui se manifeste en corpus en diachronie courte et qu'ils permettent de mettre au jour. Le chapitre 8 synthétise les descriptions menées à l'aide des indices et présente une typologie des phénomènes d'évolution caractérisés par notre approche.

La troisième et dernière partie est dédiée à la question de l'articulation entre indices et interprétation, étape fondamentale dans notre approche. En effet, le choix de mettre en place une analyse linguistique outillée place au centre de l'analyse l'interprétation des données repérées en corpus. Dans le chapitre 9, nous montrons la manière dont l'analyste doit s'appuyer sur des combinaisons d'indices pour construire une interprétation fiable et pertinente de l'évolution à partir d'indices en corpus. Le chapitre 10 nous permet d'argumenter sur le fait que cette interprétation doit être replacée au cœur d'un processus de co-construction entre indices, analyste, experts et applications. Cette dernière partie met au premier plan une question centrale mais peu décrite en langue de spécialité, à savoir la place et le rôle de l'expert dans les tâches d'analyse terminologiques.

Première partie

Pour une approche diachronique en langue de
spécialité : éléments théoriques et
méthodologiques

Chapitre 1 Diachronie et évolution dans la langue

SOMMAIRE DU CHAPITRE

1.1	PERSPECTIVE DIACHRONIQUE ET LANGUE GÉNÉRALE.....	9
1.1.1	Une « tradition » linguistique.....	10
1.1.2	Diachronie en langue générale : Trois caractéristiques	11
1.1.2.1	Diversité des niveaux de description.....	11
1.1.2.2	Facteurs internes et externes du changement sémantique.....	12
1.1.2.3	Place des questions méthodologiques	14
1.2	PERSPECTIVE DIACHRONIQUE ET LANGUE DE SPÉCIALITÉ	17
1.2.1	Fondements de la terminologie et diachronie.....	18
1.2.2	Descriptions diachroniques en langue de spécialité : bilan en demi-teinte.....	19
1.2.3	Émergence de nouvelles perspectives diachroniques	23
1.2.3.1	Nouvelles théories de la terminologie	23
1.2.3.2	Néologie, implantation et terminochronie : Trois approches diachroniques	28
1.2.3.3	Terminologie computationnelle et applications.....	35
1.3	SYNTHÈSE.....	38

Ce premier chapitre a pour objectif de situer notre recherche dans une perspective linguistique diachronique. Pour cela, nous dressons un portrait des différentes approches linguistiques centrées sur l'étude de l'évolution. Ce portrait est divisé en deux sections qui traitent respectivement de la perspective diachronique en langue générale et de la perspective diachronique en langue de spécialité. Ces deux points de vue sont ici mis en regard à la fois pour décrire la place accordée à la diachronie dans ces deux perspectives et illustrer le fait qu'aujourd'hui plusieurs éléments permettent d'affirmer que la dimension diachronique peut reprendre la place qui lui est due en terminologie. Ce rapprochement entre diachronie et langue de spécialité repose sur de nouvelles possibilités théoriques et méthodologiques mais peut également s'inspirer de l'expérience diachronique dans la langue générale pour poser les jalons d'une réflexion diachronique propre aux besoins des langues de spécialité.

1.1 Perspective diachronique et langue générale

La perspective diachronique en langue générale est, d'un point de vue historique, l'approche première de l'analyse linguistique. Si depuis Saussure elle est passée à un plan secondaire, de nombreux travaux développent toujours cette question et montrent une richesse et une maturité dans leurs réflexions dont peuvent tirer parti les langues de spécialité.

1.1.1 Une « tradition » linguistique

Au XIX^{ème} siècle, jusqu'à ce que le Cours de Linguistique Générale (CLG) de Saussure (1995 : 89-90) ne renouvelle les problématiques de la linguistique et ne développe la notion de langue comme système en synchronie, les études sur la langue adoptaient majoritairement une perspective historique. Normand (2000 : 30) rapporte à ce sujet une citation de Hermann Paul :

« La seule étude scientifique du langage est la méthode historique [...]. Toute étude linguistique scientifique qui n'est pas historique dans ses buts ni dans ses méthodes peut s'expliquer seulement soit par une déficience du chercheur, soit par l'insuffisance des sources dont il dispose. (Hermann Paul, *Principien der Sprachgeschichte*, 1880, tr. P. Caussat) ».

La dominance de la perspective historique à cette époque se retrouve également dans les fondements de la sémantique : lorsque les travaux de Darmesteter (1886)⁴ et Bréal (1897)⁵ annoncent la naissance de la sémantique comme discipline à part entière, ceux-ci la définissent à travers les notions d'*histoire* et de *transformation* :

« Jusqu'à quel point l'histoire des changements de sens reflète-t-elle l'histoire de la pensée? Tel est précisément le problème dont nous abordons ici l'examen. » (Darmesteter, 1886 : 4-5)
« Laissant de côté les changements sémantiques qui sont du ressort de la grammaire physiologique, j'étudie les causes intellectuelles qui ont présidé à la transformation de nos langues. » (Bréal, 1897 : 15)

Saussure, avant le CLG et sa proposition d'une linguistique essentiellement synchronique, publie lui-même un travail majeur en grammaire comparée où il propose une conception nouvelle du système vocalique des langues indo-européennes (Saussure (De), 1879). Ce type de travaux comparatifs doit également beaucoup aux recherches de Meillet (1905), qui développe une analyse socio-historique de la langue.

⁴ Soulignons, en cette année du bicentenaire de la naissance de Darwin et des 150 ans de la parution de son œuvre fondamentale *De l'origine des espèces par le moyen de la sélection naturelle*, que les travaux de Darmesteter s'inscrivent précisément dans une « version linguistique de l'évolutionnisme darwinien » (Nicolas, 1980).

⁵ À noter également l'article de Delesalle (1987) sur le parallèle des travaux de ces deux auteurs.

Mais dès l'introduction de son Cours de Linguistique Générale (1995 : 13^{sqq.}), Saussure cherche à extraire la linguistique de cette dimension historique. Il établit un constat de cette situation et distingue trois phases dans l'histoire de la linguistique (la grammaire en tant que discipline normative, la philologie et la grammaire comparée) desquelles il cherche à détacher la linguistique pour qu'enfin elle « reconna[isse] quel est son unique et véritable objet [la langue en elle-même et pour elle-même] » (*ibidem*).

À partir de ces travaux, l'étude de la langue se construit donc sous un jour nouveau, à travers la notion de langue comme système. L'adoption de ce principe implique d'étudier la langue en elle-même, pour ce qu'elle est, c'est-à-dire comme « un arrangement synchronique de ce qu'elle contient, et non pas une filiation diachronique de changements relatifs à son contenu » (Valin, 1990 : 89). Les études linguistiques deviennent donc avant tout des observations d'états de langue à partir desquels « le linguiste qui veut comprendre cet état doit [...] faire table rase de tout ce qui l'a produit et ignorer la diachronie » (Saussure (De), 1995 : 117).

Suite à ce renouvellement théorique, les études synchroniques sur la langue sont passées au premier plan de la linguistique moderne. Mais les approches diachroniques, bien que minoritaires, n'ont pas disparu du « paysage linguistique ». En langue générale, celles-ci restent très variées et sont soutenues par les progrès et apports de la linguistique de corpus. Cette vitalité leur permet aujourd'hui d'avancer et de mûrir dans leurs réflexions et de présenter certaines caractéristiques théoriques et méthodologiques fondamentales dont on peut s'inspirer en langue de spécialité.

1.1.2 Diachronie en langue générale : Trois caractéristiques

1.1.2.1 Diversité des niveaux de description

En langue générale, presque tous les niveaux de description linguistique intéressent les diachroniciens. La tradition linguistique des recherches en phonétique historique et en grammaire comparée se poursuit largement aujourd'hui (voir par exemple Croft, 2000 ; Hoenigswald, 1963 ; Lass, 2000 ; Traugott & Dasher, 2005), mais d'autres études diachroniques portent également :

- sur le changement sémantique (Blank, 2001 ; 2003 ; Koch, 2001 ; Rastier, 2000),
- sur l'évolution du lexique (dont essentiellement la néologie (Guilbert, 1974 ; Rey, 1976 ; Sablayrolles, 1996a ; 1996b ; 1997 ; 2002a ; 2003b)),
- sur la syntaxe (Mair, 2002 ; Mair & Leech, 2006),
- sur le phénomène de grammaticalisation (Lebel, 2003 ; Marchello-Nizia, 1995 ; 2006),
- sur des aspects pragmatiques (par exemple Sweetser, 1990),
- sur l'évolution de genres textuels (Branca-Rosoff, 1999 ; Westin & Geisler, 2002 ; Iezzi, *et al.*, 2008),
- et bien d'autres.⁶

L'intérêt de la diachronie est donc marqué en linguistique par cette diversité des niveaux de descriptions étudiés, diversité qui fait ressortir deux observations. La première est que la diachronie est marquée par différents points de vue sur les facteurs du changement linguistique, points de vue qui permettent d'introduire une distinction intéressante pour alimenter notre position en langue de spécialité entre dimension interne et dimension externe des évolutions. La seconde observation, toujours en lien avec l'objectif de cette recherche, est que toutes les descriptions proposées n'accordent pas la même place à la question de la méthodologie.

1.1.2.2 Facteurs internes et externes du changement sémantique

Les points de vue sur les causes et facteurs du changement linguistique sont divers. Pour illustrer cet aspect, on peut s'intéresser au cas des changements sémantiques, à propos desquels au moins trois types de descriptions coexistent⁷ :

- **le point de vue de la sémantique structurale** (notamment Coseriu, 1964), qui se place en tant que « développement historique des « champs conceptuels » considérés

⁶ On peut souligner que la langue des signes n'est pas en reste et porte un certain intérêt à la diachronie. Par exemple Delaporte (2005) propose une description du changement sémantique ; Bonnal-Vergès (1999 ; 2006) quant à elle observe la néologie en langue des signes.

⁷ Il existe d'autres points de vue descriptifs, peut-être moins fréquents, comme certaines études en sémantique formelle traitent la question du changement sémantique et proposent des formalisations logiques sur le modèle logique de Montague (Eckardt, 2003 dans Eckardt, *et al.*, 2003).

comme structures lexicales de contenu. Et, puisque *structure* veut dire avant tout *opposition distinctive*, la sémantique structurale diachronique aura à établir, à étudier et, si possible, à expliquer (motiver) le maintien, l'apparition et la modification, au cours de l'histoire d'une langue, des oppositions lexicales distinctives » (*ibidem* : 159-160). Cette perspective envisage donc la question de l'évolution dans les « limites » du système de la langue exclusivement ;

- **le point de vue de la sémantique cognitive** (entre autres Blank, 2001 ; Geeraerts, 1991 ; 2001 ; Koch, 2001 ; Rastier, 2000 ; 2005 ; Sweetser, 1990) : ce point de vue, plus récent, « pose que le sens linguistique consiste en représentations ou processus mentaux, ce qui [le] conduit à s'appuyer sur une psychologie ou une phénoménologie spontanée » (Rastier, 2005). L'objectif dans cette perspective est alors notamment d'essayer de mettre au jour certains universaux cognitifs sous-jacents au changement sémantique (Koch, 2001 : 80).
- **le point de vue de la stratégie « continuiste »** (Nyckees, 1998 ; 2001) qui « se fonde sur l'hypothèse que l'apparence de discontinuité que revêtent les changements de sens est d'abord un effet de la distance et des lacunes de notre information concernant les circonstances concrètes, sociohistoriques dans lesquelles ils se produisent. En conséquence, plus on parviendra à les recontextualiser, c'est-à-dire à prendre en compte le détail des circonstances sociohistoriques et discursives, plus ces discontinuités illusoires tendront à se résorber » (1998 : 76).

Ces trois points de vue offrent des éléments de réflexion quant à une distinction centrale entre *histoire interne* et *histoire externe* de la langue. En effet, si la sémantique structurale cherche les explications du changement de manière « interne », dans le système de la langue en lui-même, la sémantique cognitive s'ouvre à une dimension mentale des représentations sémantiques. La stratégie continuiste de Nyckees se place un peu plus loin dans la dimension externe puisqu'elle cherche les facteurs contextuels du changement linguistique dans les discours et les contextes socio-historiques. La prise en compte de la dimension socio-historique faisait d'ailleurs partie intégrante de bon nombre d'études historiques avant le CLG (par exemple Meillet, 1905), ce qui explique sans doute l'objectif de Saussure de

recentrer les études linguistiques sur la langue en tant qu'objet et de les mettre en garde contre une dimension trop sociologique.

Or, l'« externalisation » des analyses linguistiques n'est pas triviale, surtout lorsque l'on ose un parallèle entre langue générale et langue de spécialité. En effet, si la dimension externe n'est pas systématiquement intégrée dans les études diachroniques en langue générale⁸, celle-ci est quasiment toujours présente dans les études en langue de spécialité, ne serait-ce que par la prise en compte de la notion de *domaine*. Nous y reviendrons en détail dans la suite de ce travail (notamment au Chapitre 2), mais retenons simplement ici que l'analyse diachronique de la langue peut (voire *doit*, en langue de spécialité) s'accompagner de la prise en compte d'un parallèle certain entre point de vue *interne* et point de vue *externe*.

1.1.2.3 Place des questions méthodologiques

La dernière remarque que nous voudrions développer concerne la dimension méthodologique. Il est évident qu'en fonction des niveaux de description linguistique observés, les méthodes impliquées sont très différentes (on n'observe pas le changement phonétique comme on observe le changement sémantique). Mais au-delà de cette évidence, certaines approches se démarquent fortement par le soin qu'elles mettent à décrire leur méthodologie. C'est le cas notamment du courant de grammaticalisation (en français notamment Lebel, 2003 ; Marchello-Nizia, 1995 ; 2002 ; 2004 ; 2006 ; Prévost, 2005) et du courant de linguistique historique anglophone (par exemple Holmes, 1994 ; Mair & Hundt, 2000 ; Meurman-Solin, 1995 ; Renouf, 2002 ; Renouf & Kehoe, 2006 ; Sand & Siemund, 1992 ; Taavitsainen & Pahta, 1997). Ces approches mettent en effet au cœur de leurs préoccupations la question des corpus de textes et des outils de traitement de corpus pour l'analyse diachronique.

Les recherches sur la grammaticalisation, menées en France notamment par Marchello-Nizia et Prévost, s'intéressent au processus par lequel des mots du lexique se transforment en unités grammaticales, renouvelant ainsi la grammaire des langues. En France, ces études

⁸ Marchello-Nizia (2006 : 7) souligne justement que l'étude de l'évolution des langues connaît un renouveau impressionnant depuis deux décennies, renouveau en partie dû à l'étude de la grammaticalisation qui permet justement d'articuler l'histoire interne et l'histoire externe, souvent séparées dans la description jusque là.

s'intéressent généralement au français du XII^{ème} au XVI^{ème} siècle. Sans détailler plus avant les phénomènes de changement observés par la grammaticalisation, ce courant se démarque par la place accordée aux corpus et aux outils qui existent pour exploiter les textes d'Ancien et Moyen Français. Preuves en sont des articles récents tels que ceux de Heiden & Lavrentiev (2004), Marchello-Nizia (2002 ; 2004) ou Prévost (2005 ; 2008) qui décrivent ces problématiques et les implications théoriques que les progrès de la linguistique de corpus entraînent.

La problématique traitée dans notre travail est éloignée des descriptions de la grammaticalisation, mais au moins deux remarques largement mises en avant par ce courant doivent être soulignées, remarques dont les recherches en langue de spécialité peuvent tirer profit.

La première de ces remarques concerne le recours aux corpus numérisés et aux outils pour observer le changement. Ce point de vue de linguistique outillée met en avant le fait que les progrès de la linguistique de corpus permettent à la fois d'augmenter la taille des observables et surtout d'observer des phénomènes que l'on n'aurait pu saisir manuellement. Cet aspect est d'autant plus important en diachronie que beaucoup de phénomènes d'évolution sont très fins et difficiles à saisir. L'analyse d'états de langue anciens doit donc nécessairement être basée sur corpus. Ceci doit être mis en lien avec le second élément que nous souhaitons mettre en avant, et qui concerne la « nature de la preuve linguistique » (Marchello-Nizia, 2004 : 60) : le recours au corpus fait passer l'intuition du linguiste au second plan en permettant d'observer des données attestées. Comme le fait justement remarquer Prévost (2005 : 147) :

« Les linguistes qui s'intéressent à des états de langue contemporains peuvent choisir de recourir ou non au « corpus », comme source principale de leurs données ou comme complément de leur intuition de locuteur ou de celle de leurs informateurs. Les linguistes qui travaillent sur des états de langues disparus, en l'occurrence le français médiéval, n'ont pas ce choix : le corpus est indispensable, puisqu'il conditionne l'existence même de l'objet à décrire. »

Cependant, une caractéristique des langues de spécialité en diachronie est qu'il est particulièrement pertinent d'observer des états de langue et de connaissances spécialisées contemporains (comme nous le verrons notamment au §3.1.1.2, p.63). Il est possible alors

d'envisager une collaboration avec des experts dont on peut penser que les connaissances et l'intuition permettent de se passer d'une analyse de corpus. Néanmoins, d'un point de vue méthodologique, notre position est que la place du corpus reste centrale en diachronie, même pour des analyses contemporaines auxquelles les experts de domaine peuvent contribuer, ce que nous tâcherons de montrer dans l'ensemble de notre étude.

Cette idée selon laquelle la dimension diachronique peut se concentrer sur des états de langue contemporains est un point de vue rarement abordé en langue générale, si ce n'est par le courant baptisé ici « linguistique historique anglophone ». Ce terme renvoie aux recherches historiques sur des corpus en anglais, et notamment les recherches de l'organisation ICAME⁹ (*International Computer Archive of Modern and Medieval English*). Ces recherches abordent les questions sémantiques, grammaticales, syntaxiques et lexicologiques/lexicographiques de l'anglais moderne et de l'anglais médiéval et sont accessibles en ligne *via* le journal ICAME. Comme dans le cas du courant de grammaticalisation présenté *supra*, l'un des points forts de ces recherches repose sur le traitement et la constitution de corpus de travail. La question des ressources pour l'analyse diachronique est très bien balisée et plusieurs articles référencent les difficultés et intérêts propres au traitement des corpus diachroniques (Holmes, 1994 ; Meurman-Solin, 1995 ; Sand & Siemund, 1992 ; Taavitsainen & Pahta, 1997)¹⁰. Holmes (1994) en particulier souligne avec enthousiasme l'intérêt des comparaisons de fréquences entre corpus dans le temps et montre par là-même que la description du changement peut reposer sur l'analyse de corpus construits dans une optique diachronique :

« The prospect of using corpus data to infer language change over time is an exciting one. It is clearly possible to make suggestive and interesting comparisons between the frequencies of items in corpora of similar size and composition which have been constructed at different points in time. » (*ibidem* : 37)

Mais, contrairement à la majorité des études historiques, les recherches ICAME sur l'anglais abordent une dimension rarement explorée : celle de l'évolution de l'anglais moderne. En

⁹ <http://icame.uib.no/> (consulté le 23 février 2009).

¹⁰ Se retrouve donc en diachronie l'expertise anglophone, reconnue, sur la linguistique de corpus (cf. les travaux de J. Sinclair, D. Biber, J. Svartvik et R. Quirk, etc.).

effet, en linguistique diachronique, la question de la périodisation des données implique – du moins implicitement – que les états de langue observés sont anciens et s’étendent sur plusieurs siècles. Aborder la question du changement en anglais moderne implique de revoir ces critères pour ne plus étudier des changements révolus mais des changements contemporains souvent « en cours » (Mair, 1997). Ceci implique de fait d’adopter un point de vue dit « en diachronie courte » (Kytö, *et al.*, 2000). Cette nouvelle conception diachronique « courte » est nommée par Mair *brachychrony* (1997, repris dans Renouf, 2002) qui désigne ainsi l’étude du changement sur des intervalles courts de 10 à 30 ans.

Bien que développée en langue générale, cette notion de diachronie courte est particulièrement pertinente en langue de spécialité où les changements observés sont en lien avec l’évolution d’un domaine scientifique, évolution souvent très rapide. Nous y revenons dans le chapitre suivant.

Cette description de certains aspects saillants de l’approche diachronique en langue générale aujourd’hui permet de situer la place de la perspective diachronique dans les recherches linguistiques. Dans la mesure où la perspective diachronique est assez bien balisée en langue générale, mais assez peu en langue de spécialité, cette section a permis également de poser des pistes de réflexion dont peuvent bénéficier les langues de spécialité. Ces pistes seront développées tout au long de notre recherche, notamment la question de la prise en compte de la diachronie courte, l’articulation entre dimensions interne et externe et la construction d’une analyse outillée à partir de corpus spécialisés.

1.2 Perspective diachronique et langue de spécialité

Cette section met en regard la place de la dimension diachronique en langue de spécialité par rapport à la langue générale telle que nous venons de la décrire. La perspective diachronique en langue de spécialité est en effet une approche marginale, position qu’expliquent en partie les fondements théoriques et historiques de la terminologie (Dury & Picton, à paraître). Mais les propositions théoriques récentes, accompagnées des progrès de la linguistique de corpus, permettent aujourd’hui d’envisager une réconciliation entre langue de spécialité et diachronie, réconciliation que notre travail cherche à mettre en avant.

1.2.1 Fondements de la terminologie et diachronie

À l'inverse de la langue générale, dans le courant de la terminologie classique, il est admis que la perspective diachronique est exclue de l'analyse. Plus précisément, sans nier l'existence de la variation en diachronie (Candel, 2004), celle-ci est plutôt vue comme l'occasion d'alimenter des arguments en faveur de la nécessité de *normer* les terminologies.

Ainsi, si l'on revient aux fondements de la Théorie Générale de la Terminologie (TGT), lorsque Wüster s'efforce de dissocier la terminologie de « la science de la langue commune », l'auteur note :

« Les différences fondamentales entre la science de la langue commune et l'étude scientifique générale de la terminologie forment deux groupes, selon l'angle sous lequel on se place. Le premier groupe a trait à l'état de la langue, le second à l'évolution de la langue. » (Wüster, 1981 cité dans Rondeau & Felber, 1981 : 63)

Ce second groupe repose sur un postulat principal : celui de la formation de la langue comme activité consciente. Cette perspective permet à Wüster (*ibidem* : 63sq.) de nourrir un argumentaire sur la nécessité d'une *norme prescriptive* en terminologie, par rapport à une *norme descriptive* dans la « langue commune » :

« Jusqu'à une date récente, la linguistique n'a fait valoir que l'évolution libre, non dirigée, de la langue. C'est l'usage effectif de cette dernière qui, dans la langue commune, sert de norme. On peut appeler cette norme, la norme descriptive. En revanche, en terminologie, fertile en notions et en termes, cette évolution libre de la langue mène à une confusion inacceptable. C'est la raison pour laquelle, dès le début de notre siècle, et même plus tôt, dans le cas de quelques spécialités, les terminologues ont dû convenir d'unités de notions et de termes. Ils ont aussi donné à cette convention le nom de *norme* que l'on appelle, pour la différencier de la norme descriptive, la *norme prescriptive*. » (*ibidem* : 65)

En d'autres termes, cette observation de Wüster indique que la variation diachronique en terminologie est normée et contrôlée sur au moins deux points :

- la normalisation terminologique passe par une étape « d'unification ultérieure » (*ibidem* : 67) qui permet d'éliminer les synonymes et les homonymes potentiels « en vue d'améliorer le style ». L'évolution n'est considérée ici que dans la mesure où les variations qu'elle entraîne sont indésirables et doivent être normalisées ;
- la normalisation terminologique permet également de créer les nouveaux termes, considérant ainsi que « centraliser la création des termes assure dès le début leur unité ». L'évolution est considérée ici comme un moyen conscient de normaliser la terminologie à la source.

Bien que depuis ces travaux les études terminologiques se soient grandement développées et émancipées de la pensée wüsterienne – parfois avec virulence (Humbley, 2004), les visées normative et onomasiologique accolées aux pratiques terminologiques classiques ont été un frein certain au développement des études sur l'évolution en langue de spécialité¹¹.

Il est vrai en effet que la prise en compte de la dimension diachronique entraîne dans son champ des perspectives longtemps réfutées par la terminologie classique, telles que les changements de sens et la polysémie (Dury, 1997 ; 2000), la synonymie (Dury & Lervad, 2007) et la perméabilité entre domaines (Dury, 2008b) et entre degrés de spécialisation (Josselin-Leray, 2005 ; Meyer & Mackintosh, 2000) qui peuvent entrer en conflit avec l'idée de normalisation prescriptive. Malgré tout, ces phénomènes de variation diachronique existent et bon nombre de terminologues déplorent le manque de recherches diachroniques en langue de spécialité qui permettraient notamment de les décrire.

1.2.2 Descriptions diachroniques en langue de spécialité : bilan en demi-teinte

Nombreux sont les terminologues qui constatent que la perspective diachronique est très (et même trop) rare en langue de spécialité :

¹¹ Nous insistons sur le fait que cette conception wüsterienne de la variation diachronique se justifie clairement par la perspective de normalisation, et non par l'« ignorance » des fonctionnements linguistiques. Comme le note Humbley (1994 : 701) : « Wüster concevait sa méthode dans le cadre de la normalisation technique internationale, et même si les aspects linguistiques du travail étaient loin d'être négligés, sa visée était moins la langue que les concepts qu'il s'agissait de nommer et de définir. »

« encore trop peu de travaux terminologiques intègrent dans leur propos des considérations d'ordre diachronique. » (Dury, 1997 : 76)

Constat qui pousse certains à parler de véritable « déficit diachronique » :

« Je me suis intéressé aux répercussions de l'évolution technologique sur l'évolution des terminologies et ai dû constater que la lexicologie, la terminologie ou même la néologie, toutes disciplines traditionnelles, semblent présenter un déficit diachronique qualitatif ou quantitatif, voire les deux à la fois. » (Møller, 1998 : 426)

Ce manque pousse certains auteurs à réclamer des études diachroniques en terminologie. Il est clair en effet que l'on manque de descriptions sur l'évolution des terminologies, informations pourtant centrales pour la pratique terminologique :

« We do not know in what areas or in what elements there is the greatest change over the shortest period, nor can we reliably say what effect changes or the neglect to record these changes may have on the quality of a terminological information service. [...] Once more is known about the rate of change and life-span of terms, it may be possible to reduce this type of information ; in the meantime it is advisable to date most items of information and to plan for systematic and regular revisions of the terminological database. » (Sager, 1990 : 216)

Ce type d'information permettrait pourtant de saisir et maîtriser les phénomènes de variations terminologiques :

« Il est manifeste que le fil diachronique confère une allure plus visible à la terminologie. C'est dans cette direction qu'il faut chercher les causes des équilibres ou des déséquilibres onomasiologiques. » (Boulanger, 1995 : 198-199)

Comme mentionné *supra*, l'absence d'études diachroniques s'explique sans doute par l'influence de la terminologie classique et de ses objectifs précis de normalisation. Mais malgré cette influence et la rareté de la perspective diachronique en langue de spécialité, cette question a toutefois été ponctuellement abordée, comme l'atteste par exemple le colloque *Terminologie Diachronique* de 1988 (Schaetzen (De), 1989).

Dans la littérature diachronique en langue de spécialité, se retrouvent le plus souvent des analyses descriptives et ponctuelles de l'évolution de termes et de concepts qui intègrent une dimension historique. Celles-ci néanmoins ne sont pas toujours l'œuvre de terminologues mais le plu souvent de lexicologues¹² ou d'experts de domaines intéressés par l'histoire de leurs termes, et surtout de leurs concepts. Ces analyses, comme dans le cas d'analyses en linguistique historique¹³, privilégient les périodes anciennes et les intervalles de temps longs (Kocourek, 1982 : 23). On peut citer par exemple les travaux de Quemada (1978 ; 1955), d'Auger (1973), ou encore de Beltran-Vidal et Maniez (2005) ou de Sournia (1994), ces derniers portant sur le domaine médical, souvent étudié. On peut noter aussi certaines recherches, un peu différentes, de Banks (2003) qui décrivent non pas l'évolution terminologique mais l'évolution de la phrase en anglais scientifique.

De fait, si l'on écarte les travaux de lexicologues et d'historiens, il est vrai que les terminologues ont souvent montré un intérêt relatif à la description diachronique des terminologies. Les travaux de Dury (1997 ; 1999 ; Dury, 2007 ; 2008b) sont de ce point de vue remarquables, en particulier dans leur manière d'articuler descriptions diachroniques et réflexions théoriques en terminologie, telles que sur les notions de *concept* et de *terme* en diachronie. Ses travaux parviennent à restituer une place centrale en terminologie aux phénomènes d'emprunt, de migration entre domaines, de terminologisation et de vulgarisation (Dury, 1997 ; 2006 ; 2007 ; 2008b)¹⁴. Et si ces phénomènes sont parfois traités dans la littérature (par exemple (Delavigne, 2001a ; Meyer & Mackintosh, 2000) sur la vulgarisation, (Groult, *et al.*, 1988) sur les emprunts entre domaines, etc.), les travaux de Dury sont parmi les seuls à les problématiser nettement dans une perspective à la fois terminologique et diachronique.

¹² Comme le souligne Van Campenhoudt (1998 : 267) : « Dans l'enseignement universitaire, l'étude diachronique des vocabulaires spécialisés relève traditionnellement de la lexicologie et constitue une source inépuisable de sujets de thèses et de mémoires ; sans doute n'envisage-t-on pas assez l'intérêt d'une telle démarche pour le terminologue, réputé œuvrer en synchronie... dans l'urgence perpétuelle. »

¹³ Voir en histoire des sciences (cf. Chapitre2).

¹⁴ L'auteur parle d'*emprunt interne* pour désigner les migrations de termes entre domaines et entre degrés de spécialité, et d'*emprunt externe* pour désigner l'intégration de termes existant dans d'autres langues.

Les travaux de l’auteure font écho à une remarque amorcée *supra* : la pertinence de diachronies plus courtes pour les langues de spécialité. L’auteure souligne qu’observer l’évolution sur de longues et anciennes périodes

« entraî[n]e à la fois des constatations d’ordre étymologique, d’ordre sémantique et d’ordre terminologique. Elles font ainsi apparaître des changements sémantiques au niveau des étymons classiques, des mutations phonologiques profondes [...]. » (Dury, 1997 : 78)

Mais dans la mesure où Dury s’intéresse au jeune domaine de l’écologie, « science du 20^{ème} siècle », elle insiste sur le fait que l’on ne peut s’attendre à ce type de descriptions dans la mesure où

« les phénomènes observés [...] recourent moins systématiquement à des explications d’ordre phonétiques, phonologiques, voire même morphologiques. Ils se situent essentiellement **au niveau terminologique**¹⁵ à proprement parler. » (*ibidem*).

L’observation de périodes récentes et de diachronies courtes revêt un intérêt particulier et constitue une des spécificités de la diachronie en langue de spécialité par rapport à la langue générale. Elle permet de centrer la description « au niveau terminologique », expression qui marque une opposition – partielle au moins – avec les autres niveaux de description, centraux en diachronie en langue générale, et qui ancre la terminologie dans la dimension diachronique.

Ces remarques, qui seront développées tout au long de ce travail, montrent non seulement le besoin et la pertinence de construire des analyses diachroniques pour les langues de spécialité, mais également les particularités que celles-ci peuvent revêtir (telles que la diachronie courte ou la prise en compte de périodes contemporaines à observer), qui font de la diachronie en langue de spécialité un champ d’exploration tout à fait inédit.

Mais depuis ces vingt dernières années, la terminologie voit apparaître de nouvelles propositions théoriques et méthodologiques qui favorisent l’étude de la variation en

¹⁵ C’est nous qui soulignons.

terminologie et, par conséquent, de la diachronie. La position de la terminologie face à la perspective diachronique est influencée par ces développements, position dont la section suivante dresse le portrait.

1.2.3 Émergence de nouvelles perspectives diachroniques

1.2.3.1 Nouvelles théories de la terminologie

En réponse au caractère normalisateur de la TGT, plusieurs propositions théoriques ont émergé qui, cherchant à se démarquer des positions wüstériennes, favorisent la prise en compte des phénomènes de variation en langue de spécialité, dont la variation diachronique. On retrouve cette idée dans des propositions théoriques telles que le sociocognitivism proposé par Temmerman (2000a ; 2000b) et dont l'extrait ci-dessous illustre la position :

« En étudiant un corpus textuel de textes scientifiques adressés aux lecteurs potentiels de niveaux variés, nous avons constaté que le nombre de « vraies » notions selon la définition de « notion » proposée par l'école de Vienne est très restreint :

- 1) La plus grande partie des termes que nous avons trouvée dans notre corpus ne peut être attribuée à des notions clairement délimitées ;
- 2) Nous n'avons trouvé que quelques notions traditionnelles ;
- 3) Pour la plupart des notions, une structuration hiérarchique menant à une définition intensionnelle et extensionnelle n'est pas possible ;
- 4) Nous constatons que la synonymie (plusieurs termes pour une catégorie) et la polysémie (plusieurs catégories indiquées par le même terme) sont très présentes ;
- 5) Enfin les catégories et la signification des termes évoluent. [...]

C'est ainsi que nous nous pencherons sur *l'unité de compréhension et de la prototypicité, les modèles cognitifs, les définitions variables, le rôle de la synonymie et de la polysémie dans les processus de compréhension et l'importance de l'étude diachronique dans la compréhension, ce qui se manifeste dans le développement de modèles cognitifs métaphoriques.* » (Temmerman, 2000b : 58)

Autre proposition théorique, la socioterminologie (initiée en particulier par l'école de Rouen : Gaudin (1993 ; 2003), Guespin (1995) et Guilbert (1965 ; 1975) en qui certains voient leur précurseur), met également nettement en avant la centralité de l'évolution et de la variation en terminologie :

« Nul ne contestera certes à la terminologie spéculative le droit de rêver une langue platonicienne ou leibnizienne. [...] Pour qui [...] décide d'observer le rapport entre la terminologie comme discipline et les pratiques langagières dans des situations où fonctionnent discours scientifiques et techniques, le constat est inquiétant : une évolution sans cesse accélérée des sciences et des techniques, induisant un besoin croissant de dénominations, de transfert des savoirs et de savoir-faire, de négociation entre disciplines [...]. » (Préface de Guespin dans (Gaudin, 1993 : 9))

Et plus particulièrement sur la question de la diachronie (toujours en réaction contre la vision de Wüster) :

« La diachronie, que la théorie d'Eugen Wüster chasse par la porte, revient obstinément frapper au carreau, si fort même qu'elle ébranle la belle construction : comment expliquer sans l'histoire les constantes survivances lexicales, en terminologie comme dans le langage quotidien? » (Guespin, 1995 : 206)

Une autre approche, différente et certainement un peu moins répandue, peut être citée : celle de Kageura (1999 ; 2002) appelée « théorie de la dynamique des terminologies ». À l'inverse de la socioterminologie ou du sociocognitivism, Kageura n'insiste pas sur la variation terminologique, qu'il s'agisse de diachronie, de vulgarisation ou encore de polysémie, mais se focalise sur la question de la formation des terminologies. Plus précisément, il propose de mettre en place une théorie de la terminologie, comprise au sens de théorie des termes, qui doit nécessairement passer par une modélisation de leur dynamique :

« [...] terms manifest themselves as concrete linguistic objects within a specialised discourse and their number is constantly growing. The fact that terms are first and foremost concrete linguistic objects makes it difficult to define the theory of terms at a proper level of abstraction. [...]

In short, theories of terms – as opposed to theories of something for describing terms – are missing in the academic study of terminology. [...] Research in the field of terminology needs to be broadened to include concrete descriptive analyses of terminology based on an explicitly stated theoretical position [...]. This book is an attempt in this direction, focusing on the phenomenon of term formation and terminological growth. » (Kageura, 2002 : 1-2)

Cet extrait définit l'objectif qu'a Kageura de rendre compte de l'évolution de l'ensemble d'une terminologie (ou encore de la dynamique des termes) en modélisant la manière dont les termes se construisent les uns par rapport aux autres. En simplifiant quelque peu, on peut dire que ce point de vue propose l'étude de l'évolution des terminologies comme point de départ de la construction d'une théorie des termes. Et, bien que l'approche de Kageura ne se définisse pas comme diachronique¹⁶, elle fait la part belle à certains aspects de l'évolution terminologique, soulignant encore une fois le chemin parcouru sur cette question depuis la TGT.

Enfin, parmi les théories récentes en terminologie, une seule n'aborde pas encore la perspective diachronique : la terminologie textuelle (Bourigault & Slodzian, 1999 ; Condamines, 2000 ; Slodzian, 2000). Cette théorie s'appuie sur le lien étroit entre les pratiques terminologiques et les pratiques textuelles et font des textes la « matière première » de l'analyse :

« le texte est le point de départ de la description lexicale à construire. On va du texte vers le terme. » (Bourigault & Slodzian, 1999 : 31)

En adoptant cette conception des termes et des textes, la terminologie textuelle se rapproche de la linguistique, et notamment de la linguistique de corpus où la variation terminologique fait partie intégrante des préoccupations.

Mais également, de par leur intérêt commun pour les textes, celle-ci se rapproche des préoccupations de l'ingénierie des connaissances, ce que de nombreux articles mettent en avant¹⁷ (Bachimont, 2000 ; Bourigault & Aussenac-Gilles, 2003 ; Bourigault, *et al.*, 2004 ; Condamines, 1999 ; 2003a ; Condamines & Rebeyrolle, 1997b ; Otman, 1996 ; Rastier, 1995 ; Slodzian, 2000, etc.).

¹⁶ « [...] The concept of dynamism here is essentially synchronic, as we assume the systemic/systematic factors in the existing terminology without referring to a time scale. The dynamics of terminology addressed in the present study is, thus, the *dynamic potentiality* (of creating new terms) observed in the synchronic slice of the internal structure of terminology » (*ibidem* : 34).

¹⁷ La plupart de ces auteurs sont réunis au sein du groupe « Terminologie et Intelligence Artificielle » (TIA, <http://tia.loria.fr/presentation.html>, consulté le 3 février 2009).

De ce fait, la terminologie textuelle s'est particulièrement intéressée à la question de la construction de ressources termino-ontologiques (RTO) à partir de textes (Aussenac-Gilles, 2004 ; Aussenac-Gilles & Bourigault, 2003 ; Condamines, 1994 ; 1999 ; Otman, 1996 entre autres), « berceau » de la collaboration entre terminologie, linguistique et ingénierie des connaissances :

« En permettant d'aborder systématiquement l'étude des pratiques textuelles réelles, la linguistique de corpus, avec ses techniques et ses outils, donne accès aux expressions linguistiques concrètes d'où il sera possible de faire émerger, puis de normaliser les termes pertinents. [...]

Il va également de soi que la linguistique ne peut couvrir à elle seule le processus complet de modélisation des connaissances ; en fournissant la terminologie adéquate à l'application, le linguiste prépare le travail de représentation conceptuelle, mais il ne prend pas en charge la tâche de modélisation des connaissances qui aboutira à la construction d'une ontologie. Le relais est pris par l'ingénierie des connaissances. Le groupe TIA s'inscrit dans la nécessaire coopération interdisciplinaire entre linguistes et ingénieurs de la connaissance. » (Bourigault & Slodzian, 1999 : 32)

Ceci explique sans doute en partie pourquoi la question de la dimension diachronique ne s'est jamais posée concrètement à la terminologie textuelle : une RTO vise à représenter les connaissances d'un domaine pour une application donnée en synchronie et l'analyse terminologique a toujours été menée dans cette perspective.

Or, aujourd'hui se pose de plus en plus la question de la mise à jour et de la maintenance des RTO (Aussenac-Gilles, *et al.*, 2007a ; Aussenac-Gilles, *et al.*, 2007b), ce qui ouvre la voie vers la prise en compte de la diachronie :

« Dans la continuité d'un effort d'évaluation, il est devenu crucial de s'intéresser à un autre paramètre, laissé de côté de manière plus ou moins délibérée jusqu'à maintenant : celui de l'évolution des ressources dans le temps, c'est-à-dire de leur maintenance et de leur pertinence dans de nouveaux contextes. Le paradoxe inhérent à la constitution d'une ressource termino-ontologique devient alors flagrant : une RTO doit à la fois « normaliser » des connaissances, c'est-à-dire leur donner un statut de référence à un moment donné et pouvoir être utilisée pour accéder à des connaissances qui évoluent, parfois très rapidement, dans des contextes dynamiques. Il devient alors nécessaire de s'interroger sur l'adéquation d'une

RTO dans un contexte d'évolution dans le temps des pratiques, des textes, des connaissances et des vocabulaires. » (Aussenac-Gilles, *et al.*, 2007b)

Dans cet article, les auteures identifient quatre problématiques concernées par la question de l'évolution dans les ressources termino-ontologiques :

- la prise en compte de l'évolution dans les RTO : comment concevoir une modélisation facilement modifiable et adaptable aux évolutions? Avec quels outils?
- la relation entre les RTO et le contexte évolutif : vaut-il mieux envisager de reconstruire une RTO ou l'adapter? Comment comparer des RTO constituées à des périodes différentes?
- l'évolution du contexte : pour anticiper au mieux l'évolution d'une ressource, peut-on repérer des éléments contextuels annonçant l'inadéquation de cette ressource?
- le repérage de l'évolution dans les textes : les textes peuvent-ils fournir des indices de l'évolution des termes et des concepts?

Dans notre recherche, nous nous inscrivons dans ces nouvelles pistes théoriques en terminologie qui permettent la prise en compte de la dimension diachronique. Plus précisément, nous souhaitons nous inscrire dans une approche textuelle de la terminologie et apporter ainsi des réponses à la problématique soulevée par Aussenac-Gilles *et al.* À travers la mise en place d'une méthode d'analyse linguistique outillée sur des corpus de textes spécialisés en diachronie, nous espérons offrir des pistes pour intégrer la dimension diachronique en langue de spécialité, et en particulier en terminologie textuelle. De plus, dans la mesure où la terminologie textuelle s'intéresse particulièrement à la question de la construction des ressources termino-ontologiques, les observations apportées prennent une dimension appliquée et offrent des éléments de réflexion pour la mise à jour de ces ressources.

Aujourd'hui, en langue de spécialité, la prise en compte de la dimension diachronique et de l'évolution des terminologies se manifeste principalement à travers trois approches : l'étude de la néologie, l'implantation terminologique et la terminochronie.

1.2.3.2 Néologie, implantation et terminochronie : Trois approches diachroniques

A. Néologie

La néologie est un phénomène primordial en terminologie mais complexe et difficilement saisissable (entre autres Boulanger, 1989 ; Guilbert, 1975 ; Rey, 1976). Elle mérite un développement qui dépasse les objectifs fixés pour ce travail, mais le lecteur intéressé peut consulter des ouvrages tels que ceux de Humbley (à paraître) ou Sablayrolles (2000a).

Nous nous contenterons ici de souligner l'importance de cette notion qui, malgré les difficultés à la définir et la théoriser, est l'une des plus étudiées en langue de spécialité (pour des aperçus théoriques voir notamment Guilbert, 1974 ; 1975 ; Humbley, 2003 ; 2007 ; Rey, 1976 ; Sablayrolles, 1996a ; 2002b ; 2003b ou Rondeau, 1981 qui introduit le concept de *néonymie* ou Béciri, 2000 ; Hamelin, 1995 ; Lino, 2005, parmi d'autres, pour des analyses de la néologie dans des domaines spécifiques).

Et comme cet extrait de Guilbert l'illustre (1967 : 320 dans Guilbert, 1975 : 83)¹⁸, on ne peut parler de diachronie en langue de spécialité sans aborder la notion de néologie :

« [...] les connaissances et les techniques nouvelles ne procèdent pas de rien mais découlent de l'acquis antérieur, l'enrichissement lexical ne se réalise pas par l'introduction dans le lexique d'un ensemble homogène de mots nouveaux correspondant à chaque technique nouvelle. [...] De fait, tout ensemble lexical technique présente nécessairement dans sa genèse, un aspect diachronique. »

Les difficultés à théoriser de manière satisfaisante ce phénomène en langue de spécialité ont notamment été mises en avant par Humbley (2007), dans sa communication lors du colloque *Terminologie : approches transdisciplinaires*. L'auteur a souligné le fait que la néologie en langue de spécialité a connu un essor important dans les années 1970 et 1980, comme l'atteste la création du Réseau International de Néologie et de Terminologie (Rint) ou du Centre de Néologie et de Terminologie (CTN). Mais cet élan s'est essoufflé à la fin du siècle précédent, ce que l'auteur explique en partie par la diffusion restreinte de cette question à la

¹⁸ Dans son article de 1973b, « Théorie du néologisme » Guilbert montre la manière dont l'étude de la néologie s'ancre à la fois en synchronie et en diachronie.

francophonie, ainsi que par l'absence de théorisation satisfaisante de cette notion en langue de spécialité¹⁹.

Ceci constitue certainement une des raisons pour lesquelles la question de la néologie en langue de spécialité est souvent amalgamée à la question de la néologie en langue générale. Aussi, de nombreux travaux qui s'attachent à la définition de matrices lexicogéniques²⁰ s'adressent-ils à la fois à la langue générale et à la langue de spécialité (voir Sablayrolles, 1996b dont le recensement montre l'amalgame entre ces deux dimensions). Desmet (2002) souligne d'ailleurs que de nombreuses descriptions prouvent que d'un point de vue strictement formel, néologismes en langue générale et en langue de spécialité partagent bon nombre de caractéristiques et se répartissent en trois grands groupes (subdivisables) : la néologie sémantique, la néologie formelle et les emprunts.

Or, une des caractéristiques principales de la question de la néologie réside dans son statut d'interface entre la dimension interne (linguistique) et la dimension externe (extralinguistique) :

« Mais le lexique ne consiste pas seulement dans le système de création lexicale, il débouche sur des unités de langage liées à l'univers des choses, aux modalités de la pensée, à tout le mouvement du monde et de la société. [...] Une étude du lexique se situe à la charnière de la signification, là où opère la liaison entre les signes et les données de l'expérience. C'est un fait bien connu que le lexique se constitue par le dépôt naturel de tout l'extra-linguistique ; il s'articule sur la réalité sociale et sur son développement historique. » (Guilbert, 1965 : 8)²¹

¹⁹ Dans son ouvrage à paraître (en 2010), Humbley proposera d'ailleurs les bases d'« un programme de travaux visant à élaborer une méthodologie de la description de la néonymie, préalable à la mise au point d'une théorie, qui pourrait contribuer utilement à celle de la terminologie elle-même. » (Humbley, 2007).

²⁰ Sablayrolles (2000a ; 1996b) recense par exemple près d'une centaine de propositions de typologies des modes de création des néologismes (matrices lexicogéniques).

²¹ Il s'agit d'une idée ancienne que Guilbert (1975 : 15) reprend par exemple de Nyrop : « Qu'il s'agisse d'une découverte scientifique, d'un progrès industriel, d'une modification de la vie sociale, d'un mouvement de la pensée, d'une manière nouvelle de sentir ou de comprendre, d'un enrichissement du domaine moral, le néologisme est impérieusement demandé, et tout le monde crée des mots nouveaux, le savant aussi bien que l'ignorant, le travailleur comme le fainéant, le théoricien comme le praticien. » (Nyrop, K. (1899-1930) *Grammaire historique de la langue française*, Paris, Picard, 6 volumes).

Nous l'avons vu plus haut (§1.1.2.2, p.12) à travers l'exemple du changement sémantique, l'évolution peut être observée le long d'un parallèle entre dimension interne (linguistique) et dimension externe (extralinguistique). Dans le cas de la néologie, s'il est possible de se concentrer essentiellement sur la description linguistique interne de matrices lexicogéniques (comme c'est généralement le cas en langue générale), on peut également aborder cette question à travers le lien entre évolution des connaissances (apparition d'un nouveau concept scientifique, technique, etc.) et évolution dans la langue (création d'un nouveau terme). En langue de spécialité, où la langue est considérée comme vecteur de connaissances, la prise en compte de ce parallèle est particulièrement centrale pour l'analyse. Cet aspect peut donc être vu comme un élément spécifique des études de la néologie en langue de spécialité où il est particulièrement pertinent de s'appuyer sur ce parallèle pour décrire et analyser l'évolution (Chapitre 2).

La place centrale de la néologie est liée également aux progrès des techniques de Traitement Automatique des Langues (TAL) qui permettent depuis une dizaine d'années la mise en place d'outils de repérage automatique des néologismes. Parmi les premiers travaux à bénéficier d'une analyse automatisée, on trouve ceux de Mathieu (1998) et Mathieu *et al.* (1998) sur la langue générale, qui expliquent ainsi leur approche :

« Notre démarche consiste à confronter un corpus de journal avec le TLF en s'appuyant sur des procédures semi-automatiques. Cela permet d'alléger considérablement la tâche du linguiste, mais il reste néanmoins à ce dernier un travail important à faire : établir si un néologisme potentiel est ou non un vrai néologisme. » (Mathieu, *et al.*, 1998 : 207)

Comme l'illustrent les auteurs, ces systèmes d'extraction de néologismes font généralement appel à la notion de *corpus d'exclusion* : on compare une liste de termes existants aux termes présents dans un corpus de textes récents. L'hypothèse est que si un terme apparaît dans le corpus d'exclusion (le corpus de textes), mais est absent de la liste de référence, alors il peut s'agir d'un candidat néologisme.

Depuis ces travaux, beaucoup d'autres ont vu le jour et notamment pour les langues de spécialité. De ce point de vue, les travaux de l'Observatori de neologia de l'IULA de

l'Université Pompeu Fabra à Barcelone²² sont exemplaires (Cabré, *et al.*, 2003 ; Cabré & Yzagurri (De), 1995 ; Janssen, 2008) et développent également la question de l'extraction des néologismes oraux (Cabré, *et al.*, 2003). Parmi les développements notables de ces systèmes d'extraction de néologismes, on note qu'ils intègrent l'application de filtres morphologiques (pour l'exclusion de candidats néologismes mal construits) ou statistiques (pour ne conserver que les plus significatifs) afin d'affiner et améliorer le repérage (Roche & Bowker, 1999). La comparaison peut s'établir également directement entre textes, sans liste lexicographique de référence (Drouin, *et al.*, 2006 ; Paquin, 2007). Enfin, d'autres études développent des stratégies pour n'identifier que certains types de néologismes spécifiques, tels que les emprunts, en s'appuyant sur la comparaison d'associations de caractères et des calculs de distances entre ces associations (Minett & Wang, 2003). Ces systèmes de repérage de néologismes sont de plus en plus performants et s'instaurent désormais comme de véritables systèmes de veille néologique, notamment grâce au Web comme corpus. Si ces systèmes de veille à partir du Web sont centrés majoritairement sur l'extraction de néologismes en langue générale (Cartier & Sablayrolles, 2008), il n'est pas exclu que ceux-ci voient prochainement le jour pour les langues de spécialité.

Un peu plus loin sur l'échelle de la complexité du repérage de néologismes, se développent aujourd'hui des tentatives pour extraire automatiquement des néologismes sémantiques. La difficulté ici est que le repérage ne peut plus se baser sur le critère formel d'apparition des néologismes dans les textes, mais sur l'identification de sens nouveaux pour des termes existants. C'est la proposition par exemple de Mejri (2005). Cette proposition n'est à notre connaissance pas encore implémentée et la tâche est prometteuse mais ambitieuse.

L'importance du repérage et du traitement des néologismes se traduit aujourd'hui également par l'apparition de projets de grande ampleur tels que le projet NEOROM, coordonné par Cabré et qui réunit plusieurs observatoires de néologie des langues romanes (Cabré, 2006). Ce type de projet permet de mettre en avant les besoins applicatifs en lien avec les néologismes : « la lexicographie, la terminologie, l'harmonisation des ressources ou l'aménagement linguistique des langues [...] » (*ibidem* : 117), qui demandent une « étude permanente de l'évolution d'une langue à travers l'analyse de la néologie produite spontanément par ses locuteurs » (*ibidem* : 115). L'étude de la néologie dépasse donc le niveau descriptif et s'inscrit

²² <http://www.iula.upf.edu/obneo/obpresuk.htm> (consulté le 31 janvier 2009).

aujourd'hui dans de véritables besoins applicatifs en terminologie, caractéristique que l'on retrouve dans un autre champ de recherches en diachronie, l'implantation terminologique.

B. Implantation terminologique

Pour reprendre Auger (1999 : 18-19) cité par Gaudin (2003 : 181) :

« On peut dire que l'implantation terminologique, comme concept, est incluse dans celui de l'aménagement terminologique et qu'il n'en représente qu'un moment, si crucial soit-il. Le but recherché ultimement dans les actions d'aménagement terminologique est bien le changement linguistique et le moyen pour les États aménagistes d'agir est l'officialisation terminologique qui vise la diffusion et l'implantation de terminologies officielles. »

Le besoin identifié ici est donc d'évaluer la réussite ou l'échec de la diffusion d'une terminologie officielle. En d'autres termes, il s'agit de savoir mesurer l'impact d'une politique linguistique sur la terminologie, c'est-à-dire de comparer l'usage à la norme. Cette comparaison prend la forme d'études d'implantation telles qu'on peut en trouver par exemple chez Depecker (1997) ou Quirion (2003a ; 2003b).

Mais comme l'écrit Quirion (2003a : 7-8) :

« Comme le temps est un facteur crucial dans l'implantation de terminologies, l'évaluation ne peut prendre place que longtemps après la diffusion de celles-ci. »

De fait, l'intérêt dans le cas de l'implantation est de montrer que la perspective diachronique et le changement linguistique peuvent aller de pair avec une visée normative. La prise en compte de la variation diachronique en langue de spécialité n'est donc pas nécessairement un obstacle à la normalisation, et la description des changements se révèle être une étape nécessaire pour mieux normaliser.

L'auteur (*ibidem* : 8) ajoute :

« Ce n'est que depuis quelques années que pointent avec plus d'acuité l'importance et la nécessité de valider le résultat des efforts d'aménagement consentis. [...] Au terme du travail de diffusion terminologique, il faut pouvoir calculer précisément l'implantation des termes mis de l'avant. Pour y parvenir, un instrument d'évaluation doit être conçu. »

La mesure de l'implantation terminologique, appelée terminométrie, est donc une perspective récente qui bénéficie aujourd'hui elle aussi des apports du TAL. Ainsi, Quirion est le premier à mettre en place un protocole terminométrique automatisé. Ce protocole est implémenté dans le logiciel Barçah qui automatise « la terminométrie (comparative et diachronique) dans le but de permettre l'analyse de terminologies et de collections de documents volumineuses²³ ». Ce logiciel gère ainsi les corpus (identification de la langue, indexation), présente des concordances, désambiguïse automatiquement les termes à partir de gros corpus représentatifs, échantillonnés sur des bases statistiques pour aider l'étude.

Ce type d'approche de l'implantation est donc un exemple de perspective automatisée en diachronie en langue de spécialité, perspective récente dans le domaine. Et les travaux récents sur l'implantation se démarquent par leur besoin de mesures diachroniques scientifiques, constantes, rapides, reproductibles et outillées, travaux desquels on peut tenter de s'inspirer pour systématiser l'analyse et le repérage de l'évolution.

Dans une optique comparable de mesure (automatisée) de l'évolution des terminologies, une autre proposition a été introduite en 1998 : la terminochronie.

C. Terminochronie

Ces travaux, du fait de leur faible diffusion, peuvent prendre un caractère marginal, mais s'avèrent pourtant tout à fait intéressants dans leurs efforts pour légitimer les approches diachroniques en langue de spécialité, tout en articulant observations qualitatives et quantitatives. Ces travaux sont proposés par Møller qui introduit la notion de *terminochronie*, « l'étude de l'évolution des termes et des terminologies » (Møller, 1998 : 426).

²³ http://www.crtl.ca/fr/neuf_new.htm#barcah ou <http://www.crtl.ca/docs/barcah-final.pdf> (consultés le 17 février 2009).

L'auteur propose de distinguer macro et micro-niveaux pour la description diachronique :

- le micro-niveau renvoie au niveau du terme « isolé ». L'analyse qui y est proposée est qualitative et repose principalement sur l'observation du signifiant et du signifié des termes. L'auteur analyse notamment les processus de réduction et d'expansion des termes et présente la notion de *chaîne diachronique* qui, dans le cas d'expansions, correspondrait à une spécification du sens des termes (*disque* → *disque souple* → *disque souple double face* → *disque souple double face double densité*). Dans le cas de réductions, le terme verrait son sens conservé d'un terme à l'autre (*protection contre les rayonnements* → *protection radiologique* → *radioprotection*). Sur la base de ces observations, et sous réserve de repérer des chaînes diachroniques, l'auteur souligne l'intérêt théorique de sa proposition :

« une chaîne est donc la représentation des réductions successives de l'étendue du signifiant, repérées en discours d'abord, transformées en matrices par la suite — ou inversement : établissement de matrices théoriques d'abord, quête d'attestations par la suite. » (*ibidem* : 429)

- le macro-niveau renvoie quant à lui à l'ensemble des termes constituant une terminologie. L'analyse proposée y est quantitative. L'auteur compare l'ensemble des termes présents à la période t1 et l'ensemble des termes présents à la période t2 pour dégager :
 - A = l'ensemble des termes présents simultanément à t1 et à t2,
 - B = l'ensemble des termes qui existent à t2 mais pas à t1,
 - C = l'ensemble des termes qui existent à t1 mais pas à t2.

Son modèle lui permet de statuer sur la croissance/décroissance d'une terminologie donnée en calculant le quotient $(B + C) / A$ qui traduit le rapport entre ce qui change et ce qui ne change pas. Si celui-ci est supérieur à 1, alors la terminologie est en croissance, et inversement pour un quotient inférieur à 1.

Bien que cette description présente certaines limites²⁴, cette proposition originale soulève plusieurs questions d'intérêt. En particulier, l'optique de mesurer systématiquement l'évolution proposée par Møller met en avant le statut du terme comme moyen méthodologique d'accès à l'évolution, terme à partir duquel il est possible d'articuler mesures quantitatives globales et analyses qualitatives locales. Le point de vue de la terminochronie s'inscrit donc dans les propositions émergentes d'automatisation du repérage de l'évolution, et développe la question de l'évolution à observer au-delà de la néologie ou de l'implantation en proposant l'analyse de différents aspects tels que la disparition de termes ou l'apparition de chaînes terminologiques.

Cet article mentionne également les retombées applicatives importantes que la mise en place d'une telle approche permettrait de satisfaire :

« Un plus grand nombre d'études considérant la terminologie dans une perspective diachronique augmenteraient nos connaissances sur le potentiel évolutif des structures terminologiques, ce qui intéresserait les enseignants de langues de spécialités, les traducteurs, les lexicographes, les néologues, ceux qui normalisent la terminologie et, probablement, d'autres encore. » (Møller, 1998 : 452)

En d'autres termes, bien que les recherches de l'auteur n'aient pas été poursuivies, son article permet d'annoncer certaines des directions que prennent aujourd'hui les approches diachroniques en terminologie à travers les notions d'automatisation et d'application.

1.2.3.3 Terminologie computationnelle et applications

Aujourd'hui, avec le développement des techniques de TAL, de la prise en compte des corpus en terminologie et de l'accès aux corpus numériques, émergent de plus en plus de

²⁴ Ces limites, que l'auteur reconnaît lui-même (*ibidem* : 429), sont en particulier liées au fait que la comparaison proposée ne prend pas en compte les textes entiers mais des nomenclatures de termes, ce qui limite l'observation des phénomènes de réduction (Jacques, 2003) et entraîne des interprétations souvent rapides des phénomènes présentés. De la même manière, la modélisation de l'observation de l'évolution des terminologies sur la base du calcul $(B + C) / A$ est certainement un peu hâtive dans la mesure où l'auteur se concentre sur le nombre de signifiants recensés dans la nomenclature et non pas sur le nombre de signifiés. Ces limites s'expliquent par le fait que ces analyses se veulent les prémices d'une proposition diachronique et que les questions auxquelles celle-ci doit répondre sont vastes et complexes.

tentatives d'automatisation du repérage et de traitement des phénomènes diachroniques en terminologie. Comme décrit au paragraphe 1.2.3.2, la néologie fait figure de chef de file dans ce domaine. Mais depuis 2000, d'autres études apparaissent qui élargissent la définition d'évolution au-delà des frontières de la néologie et du repérage de nouveaux termes/concepts et qui proposent l'automatisation partielle ou complète des analyses.

C'est le cas notamment des travaux de Ahmad et son équipe (Ahmad & Al-Thubaity, 2003 ; Ahmad & Musacchio, 2004 ; Ahmad, *et al.*, 2002 ; Schierz, 2007) ou encore de Tartier (2003 ; 2004 ; 2006 ; 2007).

Les travaux d'Ahmad et son équipe s'appuient sur des méthodes et outils de linguistique de corpus pour comparer et contraster les usages linguistiques dans le temps dans un domaine et/ou un genre donné :

« A corpus-based diachronic analysis [...], based mainly on the morphologically productive use of certain terms can help in tracking the evolution of key developments in a rapidly evolving specialist field. » (Ahmad & Al-Thubaity, 2003 : 46)

« In this paper we look at key historic developments in (nuclear) physics in Italy during 1925-50 and relate these developments to modern day writings in nuclear and elementary particle physics in Italy. » (Ahmad & Musacchio, 2004 : 1567)

Les analyses sont menées à partir de gros corpus diachroniques découpés en périodes pertinentes définies en fonction de l'histoire du domaine observé (cf. 3.1.1.2, p.63).

L'un des intérêts majeurs de ces travaux pour notre perspective réside dans l'utilisation d'indices linguistiques repérables automatiquement dans les textes pour analyser l'évolution des termes du domaine. Les auteurs en utilisent principalement trois (Ahmad, *et al.*, 2002), que nous reprenons en partie et développons dans la deuxième partie de notre travail :

- les empreintes de fréquence (« Frequency Signature ») qui permettent d'observer si un terme est en croissance ou en décroissance dans les différents sous-corpus à travers le temps ;

- la productivité morphologique et lexicale, qui permet d'observer la création de termes nouveaux à partir de termes existants, par dérivation ou composition ;
- la cooccurrence, pour observer les contextes dans lesquels sont utilisés les termes (Ahmad & Musacchio, 2004).

Tartier, quant à elle, exploite dans ses travaux un parallèle entre les variantes de termes et l'évolution de ces termes. Le principe général sous-jacent à son hypothèse est que plus le mécanisme de création d'une variante est complexe, plus il est probable que le sens initial du terme ait évolué. L'auteure propose donc de calculer la distance entre un terme et sa variante pour analyser l'évolution. Pour prendre un exemple simple de distance, l'auteure considère qu'il faut trois modifications pour passer du terme *diffusion cohérente inélastique d'électrons* à sa variante *diffusion cohérente des neutrons thermiques* : 1- suppression (*diffusion cohérente d'électrons*), 2- substitution (*diffusion cohérente des neutrons*), 3- insertion (*diffusion cohérente des neutrons thermiques*) (Tartier, 2006)²⁵.

Force est de constater cependant que ces études sont assez souvent très informatiques et mathématiques, et ne questionnent pas réellement leurs objets d'un point de vue terminologique ou linguistique²⁶. Elles apportent néanmoins plusieurs pistes intéressantes pour automatiser l'analyse dans une optique linguistique et terminologique. Notamment, les indices proposés peuvent être repris et discutés pour interroger le parallèle possible entre indices linguistiques dans les textes et l'évolution des connaissances, démarche proposée dans notre recherche.

Ces approches s'inscrivent dans la lignée de nouvelles possibilités d'automatisation des analyses et descriptions diachroniques en corpus. Les possibilités de traitement sont plus grandes, mieux balisées et permettent de répondre maintenant à différents besoins et applications pour lesquels la diachronie est centrale mais auxquels on ne pouvait répondre faute de moyens. Parmi ces besoins, nous avons cité dans ce chapitre la mise à jour de ressources terminographiques (Aussenac-Gilles, *et al.*, 2007b), la veille néologique (Cabré, *et*

²⁵ Cette approche interprète donc « à l'extrême » le parallèle entre dimension interne (linguistique, la dénomination) et externe (extralinguistique, ici le concept). Nous reviendrons sur cette question.

²⁶ D'où l'impression que le linguiste peut avoir parfois de la rapidité et de la simplification des interprétations.

al., 2003 ; Sablayrolles, 2002b ; Cartier & Sablayrolles, 2008) et les études d'implantation terminologique (Quirion, 2003a ; Quirion, 2004), mais d'autres applications sont envisageables. Ainsi, la veille scientifique et technique (Ibekwe-Sanjuan, 2005), la traduction (Dury, 2002) ou de la didactique des langues de spécialité (Celotti & Musacchio, 2004 ; Ravaoarimalala, 2005) manifestent un intérêt marqué pour ces questions²⁷. Ces perspectives et les nouveaux apports de l'analyse diachronique participeront sans aucun doute ces prochaines années au développement de la prise en compte de la dimension diachronique pour les langues de spécialité et permettront en retour de mieux connaître les phénomènes d'évolution dans les langues et domaines de spécialité.

1.3 Synthèse

Nous avons choisi de construire ce premier chapitre comme un portrait croisé de la place de la diachronie en langue générale et en langue de spécialité. Le bref retour historique sur la tradition diachronique en langue générale a eu pour but d'inscrire notre perspective dans une orientation linguistique, mais surtout de montrer la manière dont les langues de spécialité se sont clairement distinguées de la langue générale sur la question de la diachronie. Au-delà de ce constat historique, ce portrait argumente en faveur du fait qu'aujourd'hui, tous les éléments sont réunis pour réconcilier enfin les langues de spécialité avec la perspective diachronique, très peu explorée jusque là.

Cette réconciliation peut se construire aujourd'hui sous au moins deux angles complémentaires : un angle théorique et un angle méthodologique (Dury & Picton, à paraître).

L'angle théorique est soutenu par la mise en place depuis une vingtaine d'années de nouvelles propositions théoriques en terminologie qui argumentent en faveur de la prise en compte de la variation linguistique. La terminologie dispose donc aujourd'hui de nouvelles bases théoriques sur lesquelles asseoir une description et un ancrage fort des phénomènes diachroniques dans son champ, dont la terminologie textuelle.

²⁷ Notons également que dans notre étude, les experts qui ont collaboré avec nous ont manifesté un intérêt certain pour notre approche dans une perspective d'aide à la rédaction de leur documentation, des cours de référence, de la mise à jour des différentes éditions de cours, etc.

Le deuxième angle d'approche pour nourrir cette réflexion est un angle méthodologique. En effet, malgré les besoins reconnus de descriptions diachroniques des langues de spécialité, celles-ci ne sont pas faciles à mettre en œuvre et seuls les progrès récents de la linguistique de corpus et du Traitement Automatique des Langues permettent aujourd'hui d'envisager de systématiser, au moins en partie, les descriptions. C'est donc par la mise en place de méthodes d'analyse de l'évolution que l'on pourra disposer de données et de retours suffisants pour ancrer la dimension diachronique dans les langues de spécialité. Notre recherche s'inscrit dans cette perspective, ce qui justifie également notre choix de nous lier à la terminologie textuelle, qui place au centre de ses préoccupations la mise en place de méthodes linguistiques outillées pour les analyses terminologiques.

Notons qu'un troisième élément favorise cette réconciliation : la perspective applicative. En effet, les analyses diachroniques n'ont pas pour seul objectif de décrire les phénomènes d'évolution, mais sont également la clé de nombreux enjeux applicatifs sur lesquels nous reviendrons dans ce travail, et en particulier dans le Chapitre 10. Le développement de la perspective diachronique est loin d'être une question marginale ou infondée, mais se présente comme un véritable besoin et un véritable enjeu pour la terminologie.

Pour atteindre ces différents objectifs, et pour que les langues de spécialité s'approprient la perspective diachronique, il est important que les analyses n'ignorent pas l'expérience de la langue générale. Si les relations entre langue générale et langue de spécialité se sont traditionnellement construites sur la base d'une opposition, les langues de spécialité peuvent chercher à bénéficier de l'expérience et des réflexions mûries des études en langue générale pour construire un objet qui leur est propre. Plutôt que d'ignorer les travaux proposés en langue générale, ces derniers mettent au contraire au jour certains éléments de réflexion dont les langues de spécialité peuvent s'inspirer pour construire leur propre approche diachronique. Parmi ces aspects, ceux que nous retenons et développons dans les chapitres suivants sont principalement la place primordiale accordée aux corpus pour l'analyse diachronique, les nouvelles possibilités d'outillage pour l'analyse et la distinction entre une linguistique interne et une linguistique externe.

C'est d'ailleurs à ce dernier parallèle que nous faisons appel dans le chapitre suivant. En effet, cette distinction nous sert de fil conducteur pour présenter les principales notions théoriques en jeu et fonder les bases d'une approche méthodologique linguistique et diachronique en langue de spécialité.

Chapitre 2 Pour un parallèle entre évolution de la langue et évolution des connaissances

SOMMAIRE DU CHAPITRE

2.1	INTERNE/EXTERNE : UNE DISTINCTION THÉORIQUE FONDAMENTALE EN TERMINOLOGIE	42
2.1.1	Définition du terme	42
2.1.1.1	Du « concept et terme-étiquette »	42
2.1.1.2	... au « terme-signe linguistique et concept »	44
2.1.1.3	Appartenance à un domaine	45
2.1.2	Nouveaux rapports entre textes et connaissances	48
2.1.2.1	Terminologie textuelle	48
2.1.2.2	Dimension externe : définir les « connaissances »	51
2.1.2.3	Dimension interne : la langue et les textes	54
2.2	POINTS DE VUE HISTORIQUE ET DIACHRONIQUE	55
2.2.1	Histoire interne et histoire externe	55
2.2.2	Hypothèse d'un parallèle entre langue et connaissances pour la diachronie	59

L'hypothèse de notre recherche se fonde sur un parallèle entre évolution dans la langue et évolution des connaissances. Dans ce chapitre, nous présentons les arguments théoriques qui permettent de poser ce parallèle en langue de spécialité et en diachronie. Pour cela, nous nous appuyons sur la distinction qu'opère la linguistique structurale entre dimension *interne* et dimension *externe*. Cette distinction met en opposition une linguistique *interne*, centrée sur la langue comme système, et une linguistique *externe* qui s'intéresse à « tous les points par lesquels la linguistique touche à l'ethnologie, toutes les relations qui peuvent exister entre l'histoire d'une langue et celle d'une race ou d'une civilisation » (Saussure (De), 1995 : 40-43), mais aussi à l'histoire politique, aux rapports entre langues et institutions (écoles, église, etc.) et à l'extension géographique des langues (Helgorsky, 1981). Cette distinction intervient dans les fondements théoriques et méthodologiques de la terminologie qui s'en inspire pour définir son objet d'étude. Retracer la position des langues de spécialité en regard de cette distinction permet ainsi de présenter l'ancrage et les enjeux théoriques de notre approche.

Dans la première section, nous montrons que cette distinction est au centre des changements théoriques et méthodologiques connus par la terminologie depuis quelques années et qu'elle permet de poser un lien entre langue et connaissances dans les langues de spécialité. Dans la deuxième section, nous illustrons la manière dont cette distinction traduit différents points de

vue possibles en diachronie et précisons la position choisie pour mettre en œuvre un parallèle entre évolution dans la langue et évolution des connaissances dans notre étude.

2.1 Interne/externe : une distinction théorique fondamentale en terminologie

L'objectif de cette section est de montrer que la distinction entre ce qui est interne (linguistique) et externe (extralinguistique) se retrouve au cœur des théories de la terminologie et permet de mieux comprendre les enjeux des changements théoriques observés dans la discipline ces dernières années.

Cette distinction est notamment essentielle pour définir le terme, étape qui, comme le note Josselin-Leray (2005 : 7) « a déjà fait couler beaucoup d'encre ». Dans son article de 2001, Petit cherche d'ailleurs à souligner la difficulté à saisir « l'identité du terme » et montre la manière dont la terminologie s'est basée sur la linguistique pour mieux s'en dissocier et décrire son objet et ses objectifs propres.

2.1.1 Définition du terme

2.1.1.1 Du « concept et terme-étiquette »...

Dans la définition classique du terme, la Théorie Générale de la Terminologie présente son objet d'étude comme étant avant tout le concept. L'adoption de cette vision cognitive permet précisément à la terminologie de s'opposer à la conception linguistique (Petit, 2001 : 63-64) en s'inscrivant avant tout dans la dimension externe (extralinguistique). De fait, le terme est quant à lui réduit à sa plus simple expression linguistique :

« considéré comme l'étiquette apposée sur une 'unité de pensée' qu'il désigne, [...] [il] se distinguerait ainsi du signe saussurien dont les deux faces [signifiant et signifié] relèvent de l'univers linguistique. » (Thoiron & Béjoint, 2003 : 3-4)

L'objectif du terminologue dans cette perspective est avant tout de sélectionner, décrire et organiser les concepts d'un domaine et, dans un second temps, de leur accoler une étiquette linguistique :

« En terminologie [...] la question est non pas de savoir ce que signifie telle forme linguistique, mais bien plutôt, une notion ayant été délimitée clairement, de savoir quelle est la forme linguistique qui la représente. Autrement dit, à l'opposé de la démarche du lexicologue, qui, ayant identifié une forme linguistique, cherche à établir le ou les sens qui s'y rattachent, le terminologue part d'un concept (signifié) pour se demander comment il se nomme ou, à défaut, comment ce concept pourrait se nommer. » (Rondeau & Felber, 1981 : 21)

Cette vision minimale de la place de la linguistique est accentuée par le principe de bi-univocité. Ce principe renvoie à deux critères : la monoréférencialité et l'univocité. L'univocité signifie que, pour un terme donné, « à un [concept]²⁸ correspond une dénomination donnée et une seule ». Le principe de monoréférencialité est le fait que pour un terme donné « à une dénomination correspond un [concept] et un seul » (Rondeau, 1981 : 24). Ces deux critères impliquent donc que le terme ne peut être que monosémique et n'avoir aucun synonyme.

Comme le notent de nombreux auteurs (parmi lesquels Depecker, 2002 : 40 ; Desmet, 2007 ; Gaudin, 2003 : 44 ; Josselin-Leray, 2005 : 9 ; Thoiron & Béjoint, 2003 : 3), cette vision du terme ne s'applique que dans le cadre de terminologies extrêmement normalisées dans la mesure où la bi-univocité prévient tout type de variation. Or, les perspectives descriptives montrent toutes que le terme est sujet à variation lorsque l'on observe son fonctionnement en discours, qu'il s'agisse de polysémie, de synonymie, de variantes, etc. (Bouveret, 1998 ; Collet, 1997 ; Condamines & Rebeyrolle, 1997a ; Dury, 1997 ; L'Homme, 2004 et bien d'autres).

D'un point de vue diachronique, adopter une définition du terme comme étiquette d'un concept préexistant et exempte de toute variation reviendrait à réduire la description de l'évolution aux cas de création de concepts par des experts du domaine auxquels ils attribueraient une « étiquette ». Cela impliquerait donc que l'évolution des terminologies serait un phénomène dont les experts ont nécessairement conscience et qui se résumerait à l'innovation et à la créativité. Or, et notre travail en fournira de nombreux exemples, cette vision est inconcevable en diachronie.

²⁸ L'auteur utilise le terme *notion* plutôt que *concept*. Dury (1997 : 23sq.) détaille la position de plusieurs auteurs qui proposent une distinction entre ces deux termes (dont Gentilhomme, 1994 ou Rey, 1979). Pour notre part, nous considérons ces deux termes comme synonymes (à l'instar de Dury), mais réservons *concept* aux emplois associés à la terminologie et aux langues de spécialité.

Pour ces différentes raisons, la définition classique du terme, très réductrice d'un point de vue linguistique, a été beaucoup rediscutée ces dernières années (L'Homme, 2005 ; Cabré, 2003), discussions qui ont mené à un rééquilibrage nécessaire entre dimension interne et dimension externe pour l'analyse terminologique.

2.1.1.2 ... au « terme-signe linguistique et concept »

Au vu de l'inadéquation du modèle classique face aux fonctionnements observés en corpus, de nombreux auteurs insistent sur le fait que « les cloisons entre la terminologie et la linguistique ne se justifient plus » (L'Homme, 2000 : 72). Plutôt que de s'opposer à la linguistique en minimisant la dimension interne du terme, les nouvelles propositions théoriques aujourd'hui suggèrent donc un rapprochement entre ces deux disciplines et considèrent le terme comme signe linguistique (Chapitre 1).

On assiste donc à un « rééquilibrage » entre linguistique et extralinguistique, étape nécessaire pour pouvoir rendre compte du véritable fonctionnement des termes en discours et décrire la variation, dont la variation diachronique. Il en ressort alors que le parallèle linguistique/extralinguistique n'est plus envisagé comme une *opposition* (vision saussurienne) mais comme une *association*²⁹. Cette idée d'« association » n'est pas triviale et met au jour le lien essentiel entre linguistique et extralinguistique dans les langues de spécialité. Comme le souligne Rey (1988 cité dans Gaudin, 1993 : 119) :

« Dans son histoire, la terminologie s'oppose à la description des langues. Loin de constater et d'analyser un fonctionnement lié à une communauté linguistique, elle cherche à régler un autre fonctionnement, qui est lié à l'exercice des connaissances. »

²⁹ Néanmoins, comme le note entre autres Petit (2001), ces nouvelles propositions ne résolvent pas toutes les questions quant à « l'identité du terme ». En particulier, beaucoup de discussions existent encore sur le rapport entre signe et concept (Depecker, 2002 ; Petit, 2001) et donnent lieu à plusieurs propositions. À titre d'illustration, dès 1981, Rondeau envisage le terme comme signe linguistique, composé d'un signifiant et d'un signifié, et assimile ce dernier au concept. Lerat (1989) introduit lui aussi une conception du terme comme signe linguistique mais propose cependant de bien distinguer le signifié du concept en adossant plutôt le concept, externe à la langue, au couple signifié/signifiant, interne. On retrouve une vision comparable chez Depecker (2000 ; 2002).

La spécificité du terme repose donc sur sa « double nature » (Lerat, 1995) ou son « double fonctionnement » (Condamines, 1994 ; 2003a)³⁰ : il est à la fois signe linguistique, comme n'importe quel mot de la langue en même temps qu'il constitue la dénomination d'un concept. De fait, toutes les langues de spécialité sont « tout à fait du français, et en même temps [...] le vecteur de savoirs et de savoir-faire » (Lerat, 1995 : 11-12).

Lerat (*ibidem* : 23) note que cette « double nature des termes³¹ brouille les frontières saussuriennes entre linguistique immanente et linguistique externe ». Plus précisément, plutôt que de « brouiller » les frontières entre linguistique interne et linguistique externe, les langues de spécialité permettent une association entre ces deux dimensions qui leur est propre, association sur laquelle l'analyse terminologique peut s'appuyer pour les descriptions. C'est ce que propose notamment la terminologie textuelle qui exploite de nouveaux rapports entre textes et connaissances pour la construction de ressources terminologiques.

Mais avant de détailler cela plus avant, nous nous arrêtons sur un dernier élément central pour la définition du terme : le *domaine*. Cette notion fonde l'une des différences essentielles entre langue générale et langue de spécialité mettant en jeu l'importance de la dimension externe pour l'étude de ces dernières.

2.1.1.3 Appartenance à un domaine

Pour circonscrire les connaissances transmises en langue de spécialité, celles-ci doivent être définies à travers la notion de domaine. Cette notion est en effet centrale pour définir les termes, comme le souligne Gaudin (2003 : 50) : « l'appartenance à un domaine figure parmi les critères centraux qui permettent traditionnellement de définir le terme ». Lerat (1995 : 16) également explique que le terme est « à la fois un signe linguistique et un contenu de connaissance, une notion (ou concept) qui se définit explicitement **en référence à un domaine**

³⁰ L'auteure ajoute que « presque tous les types d'utilisation de la terminologie dans des projets applicatifs font appel à des ressources qui doivent rendre compte du double fonctionnement du terme : comme signe linguistique et comme clé d'accès à une connaissance spécialisée » (Condamines, 1994 : 34).

³¹ Comme le note (Slodzian, 2000 : 66) « le mot « terme » sert à désigner à la fois l'unité terminologique complète (dénomination + concept) en même temps qu'il fonctionne comme synonyme de dénomination ». Désormais, dans la suite de ce mémoire « terme » sera utilisé essentiellement pour renvoyer à la « dénomination » et « terme/concept » servira à désigner l'unité terminologique complète.

du savoir dont il est solidaire³² ». L'importance du domaine pour définir la notion de *terme* est telle que certaines propositions font reposer la distinction entre *terme* et *mot* précisément sur l'appartenance du premier à un domaine. Par exemple, Cabré (2000 : 34-35) écrit que « ces unités ne sont pas des termes ou des mots en soi, mais elles activent leur caractère de mot ou de terme en fonction des caractéristiques pragmatiques de la situation dans laquelle elles sont utilisées ».

Plusieurs définitions de *domaine* ont été proposées. Parmi celles-ci, pour De Bessé (2000 : 183) « un domaine est une structuration de connaissances ». Ou encore, Poibeau (2005 : 4), écrit que « [d]'un certain point de vue [...], il s'agit d'une notion empirique qui reflète le champ de savoir d'un expert (on parle à ce propos d'*expert du domaine*) ».

Il ressort des différentes définitions proposées que deux visions coexistent quant à la délimitation des domaines :

- soit on réunit un ensemble de concepts qui forment alors un domaine (c'est ce que Sager (1990 : 37) nomme une démarche « bottom-up »),
- soit l'on décide *a priori* d'un découpage en domaines pour lesquels il s'agira de retrouver et de structurer les concepts (démarche « top-down » (*ibidem*)).

Comme le souligne Delavigne (2001a : 167), dans les faits, c'est la seconde approche qui est généralement adoptée et les domaines naissent d'un découpage préalable. Mais l'on se confronte alors à une telle multiplicité de domaines qu'il faut une « âme de collectionneur et une persévérance à toute épreuve » pour les lister (Delavigne, 2001a : 167 cité par Josselin-Leray, 2005 : 35)³³.

Cette difficulté à délimiter les domaines repose également sur la nécessité de les distinguer en fonction de leur nature (Bessé (De), 2000 ; Depecker, 2002). Dans notre recherche, nous ne reprenons qu'une seule distinction proposée dans la littérature et qui conditionne une partie de nos observations : la distinction entre domaine de connaissance et domaine d'activité. De Bessé (*ibidem* : 184) définit un domaine de connaissance comme « un savoir constitué, structuré, systématisé selon une thématique ». Il définit un domaine d'activité comme

³² C'est nous qui soulignons.

³³ Josselin-Leray (2005 : 35-36) rapporte plusieurs enquêtes qui illustrent la multiplicité et la difficulté à dénombrer et délimiter les domaines.

« permet[tant] d'identifier un champ d'action, un ensemble d'actes coordonnés, une activité réglée, une pratique », domaine qui « correspond à une activité humaine, sociale, économique, [ou] industrielle » et « est constitué d'un ensemble de procédés bien définis destinés à produire certains résultats ». Cette distinction sera reprise notamment dans le Chapitre 3 (§3.2.1.2) car elle montre l'importance de tenir compte des types de domaines pour observer l'évolution.

Ces différentes réflexions sur la difficulté à délimiter un domaine pousse certains auteurs, tels que L'Homme *et al.* (2003 : 153) à envisager que « la définition d'un « domaine » est arbitraire et relative aux objectifs définis pour un projet terminologique », réflexion que nous adoptons. En effet, aussi difficile qu'il soit de les définir et de les délimiter, ce qui importe en diachronie est surtout de ne pas considérer les domaines comme des « entités » fixes hermétiques. À la suite de Delavigne (2001a : 169), nous nous accordons sur le fait que la notion de domaine « doit [...] se penser à la fois dans son hétérogénéité externe (multiplicité des disciplines qui traverse[nt] un domaine donné) et dans son hétérogénéité interne (multiplicité des points de vue au sein dudit domaine) ». Autrement dit, quel que soit le domaine défini pour une recherche en diachronie, il faut nécessairement accepter le fait qu'un domaine n'est pas un objet homogène. Les termes/concepts « contenus » dans ce domaine peuvent migrer vers un autre domaine ou à l'inverse un domaine peut verser de nouveaux termes dans la terminologie d'un autre (Dury, 1997). Il faut également envisager qu'un domaine, puisqu'il correspond à une structuration de connaissances, puisse évoluer lui-même. Enfin, il faut accepter que différents points de vue peuvent coexister dans un domaine donné³⁴, ce que la terminologie textuelle met clairement en avant :

« Étant donné un domaine d'activités, il n'y a pas une terminologie, qui représenterait le savoir sur un domaine, mais autant de ressources terminologiques que d'applications dans lesquelles ces ressources sont utilisées. [...] On est loin de la conception idéalisée du domaine comme fragment de connaissances bien structurées, permanentes et clairement circonscrites. » (Bourigault & Slodzian, 1999 : 30-31)

³⁴ Ce que décrivent des analyses telles que celle de Condamines et Rebeyrolle (1997a).

Pour ces différentes raisons, il apparaît clairement que la délimitation d'un domaine est décidée à un moment précis pour un objectif précis mais qu'il s'agit d'un « univers » mouvant.

La prise en compte du domaine pour définir le terme rejoint alors l'importance de la prise en compte de la dimension externe pour définir l'objet d'étude des langues de spécialité. Cela permet également de souligner le fait que dimensions interne et externe ne se résument pas au terme et au concept. Bien que cette association soit centrale, il faut l'envisager de manière plus globale pour comprendre les enjeux de l'analyse diachronique des langues de spécialité. Plus généralement, les nouvelles propositions théoriques et le rééquilibrage entre linguistique et extralinguistique pour l'étude des langues de spécialité forcent à s'interroger sur les rapports nouveaux entre textes et connaissances et non plus seulement entre terme et concept. De ce point de vue, la terminologie textuelle offre de nombreuses pistes de réflexion dont nous tirons parti pour l'analyse diachronique.

2.1.2 Nouveaux rapports entre textes et connaissances

2.1.2.1 Terminologie textuelle

Le fait de considérer le terme comme signe linguistique à part entière entraîne de lier la pratique terminologique aux données textuelles. En effet, comme le souligne Kocourek (1991a : 33), les textes spécialisés :

« constituent la liaison interdisciplinaire permanente entre la spécialité et la linguistique. La langue de spécialité, textes et système, est, d'une part, l'instrument de la connaissance spécialisée – elle fait donc partie de la spécialité et devrait être citée ensemble avec les autres composantes de la spécialité ; elle est, d'autre part, le sujet de l'analyse linguistique. »

L'observation des langues de spécialité doit donc partir des textes « produits ou utilisés par une communauté d'experts » dans la mesure où c'est là « que sont exprimées, et donc accessibles, une bonne partie des connaissances partagées de cette communauté » (Bourigault & Slodzian, 1999 : 30). Les corpus de textes permettent alors de saisir les termes en contexte et

de mettre en place une réelle approche descriptive de leur fonctionnement linguistique pour accéder aux connaissances du domaine.

Cette proposition, appelée « terminologie textuelle », bouleverse l'analyse terminologique classique sur au moins deux plans.

Tout d'abord, elle implique de ne plus penser l'analyse comme onomasiologique mais comme sémasiologique. Cette distinction renvoie au fait que dans l'optique classique, le concept étant l'objet premier de la terminologie, l'analyse prend pour point de départ le concept et lui attribue un terme-étiquette. Cette démarche est appelée démarche *onomasiologique*. À l'inverse, la terminologie textuelle impose de partir non pas du concept mais des termes dans les textes pour accéder au concept. Cette démarche est appelée *sémasiologique*. En diachronie, certaines études se rapprochent d'une démarche onomasiologique et font du concept leur point de départ. Leur objectif alors est de décrire l'évolution d'un concept et de sa dénomination au fil du temps. C'est le cas par exemple de Dury (1997 ; 1999) à propos de concepts de l'écologie, Hamelin (1995) à propos du concept *nordicité*, Humbley (1994) sur le concept de *phonographe*, ou encore Van Campenhoudt (1998) sur les concepts *maille* et *maillon*. Dans les faits cependant, la démarche onomasiologique est souvent une « gageure » (Thoiron & Béjoint, 2003 : 7). Ainsi, Sager (1990 : 56, cité par Josselin-Leray (2005 : 30)) insiste sur le fait que :

« In reality the onomasiological approach only characterises the scientist who has to find a name for a new concept (an invention, a new tool, measurement, etc.); the terminologist, like the lexicographer, usually has an existing body of terms to start with. Only rarely is a terminologist involved in the process of naming an original concept. »

D'un point de vue diachronique, cette démarche ne nous semble pas adaptée sur plusieurs plans. Dans notre recherche notamment, il ne s'agit pas de décrire l'évolution de concepts connus *a priori* (élément central si l'on souhaite entreprendre une démarche onomasiologique), mais d'identifier les concepts susceptibles d'avoir évolué. Cet aspect est d'autant plus important qu'il est clair que l'évolution des connaissances est un phénomène dont les experts n'ont pas toujours conscience et pour lequel l'analyste (linguiste ou

terminologie) n'a pas d'intuition. La seule démarche possible est alors de partir des textes pour repérer les connaissances évolutives et d'adopter la démarche sémasiologique décrite en terminologie textuelle³⁵.

Le second aspect que nous souhaitons ici mettre en évidence est le fait qu'adopter une démarche textuelle, sémasiologique, implique de considérer le concept comme un construit et non pas comme une entité préexistante. Cette vision se retrouve chez les socioterminologues qui considèrent que le concept se construit en discours et qu'il s'y modifie en fonction de variables sociales et historiques (Gaudin, 2003). Elle est reprise en terminologie textuelle et notamment par Condamines (2003a : 52-53). L'auteure, à la suite des travaux de Rastier (1995), argumente en faveur d'une définition du concept comme « signifié normé » et pose que le concept ne peut être établi que si

« – il existe des conditions d'énonciation communes à un ensemble de locuteurs qui permettent de neutraliser les éléments propres à ce locuteur ; on a pu parler d'un locuteur collectif dans le courant de l'analyse de discours, ou de communautés de locuteurs (Gaudin, 1995) ;
- un interprétant (terminologue, linguiste, documentaliste, expert...), ayant un objectif précis, et qui, à partir du constat de régularités « immanentes » (normaison), attribue à certains des signifiés d'un texte, le statut de concept (normalisation) ; il y a donc passage d'un système sémiotique à un autre, relevant tous les deux du linguistique. » (*ibidem*)

Cette vision du concept comme signifié normé ouvre des perspectives centrales pour aborder l'analyse diachronique. En effet, considérer que le concept est construit implique de reconnaître que :

- le lien entre la forme linguistique analysée dans les textes et les connaissances n'est ni direct, ni univoque. Il doit être établi après une étape d'interprétation où l'analyste tient compte de régularités linguistiques observables en corpus ;
- l'interprétation doit faire intervenir des experts du domaine qui jugent de la validité et du statut des informations extraites ;

³⁵ La démarche sémasiologique est également la source d'un rapprochement entre terminologie et linguistique de corpus, rapprochement qui ouvre la voie à l'utilisation d'outils d'exploration (concordanciers, extracteurs de candidats termes, extracteurs de relations, etc.). Cette question est détaillée au chapitre suivant.

- l'interprétation doit prendre en compte les types de textes qui constituent les corpus d'étude, ainsi que l'objectif pour lequel l'étude est menée (Condamines, 2003a : 53).

Accepter que le lien entre dimension interne et dimension externe est construit est primordial lorsque l'on envisage ce lien au niveau des notions de *langue* et de *connaissances*, et non plus au niveau du terme et du concept seuls. Nous posons ces éléments comme postulats à ce stade, mais justifions dans l'ensemble de notre travail de leur pertinence en diachronie.

Pour mieux comprendre la pertinence de ce point de vue, il convient d'éclaircir ce que peuvent recouvrir les dimensions de *langue* et de *connaissances* lorsque l'on travaille à partir de textes.

2.1.2.2 Dimension externe : définir les « connaissances »

Nous l'avons vu, les langues de spécialités sont définies à la fois comme des langues et comme des vecteurs de connaissances. Cette notion de *connaissances*, que nous considérons comme relevant de la dimension externe, fait cependant rarement l'objet d'une définition, bien qu'elle renvoie à des éléments hétérogènes.

Généralement, la notion de *connaissances* renvoie avant tout au *concept*, « représentation mentale qui retient les caractéristiques communes à un ensemble d'objets » (L'Homme, 2004 : 11) ou « unité structurée de pensée par laquelle nous nous formons une connaissance du réel » (Depecker, 2002 : 43). Dans le cadre de la construction de ressources terminologiques, le concept est souvent complété par la notion de *relations conceptuelles*. En effet, ce qui intéresse le terminologue ce sont non seulement les concepts, mais également les relations qu'ils entretiennent entre eux, permettant de les structurer en *réseaux de connaissances*³⁶. Cet aspect illustre une fois encore le fait que d'un point de vue externe, le domaine est un élément central à prendre en compte.

De manière générale, la dimension extralinguistique renvoie à des réalités diverses qu'il n'est pas toujours facile de délimiter. Ceci est accentué par le fait que depuis quelques années, l'émergence d'une terminologie textuelle s'accompagne de nombreux changements dans les

³⁶ Rappelons qu'un domaine est défini comme étant « une structuration de connaissances » (Bessé (De), 2000 : 183), ce qui implique bien des concepts et les relations qui les lient et les organisent.

pratiques et applications concernées par les analyses terminologiques³⁷ (Bourigault & Aussenac-Gilles, 2003). À chacune de ces applications correspondent des types de ressources terminologiques différents construits à partir de textes³⁸. Ainsi, Bourigault et Aussenac-Gilles (*ibidem* : 2) recensent entre autres des ressources telles que :

- des thesaurus pour les systèmes d'indexation automatique ou assistée,
- des index hypertextuels pour les documentations techniques,
- des terminologies de référence pour les systèmes d'aide à la rédaction,
- des référentiels terminologiques pour les systèmes de gestion de données techniques,
- des ontologies pour les mémoires d'entreprise, pour les systèmes d'aide à la décision ou pour les systèmes d'extraction d'information,
- des glossaires de référence et listes de termes pour les outils de communication interne et externe.

Ces différentes ressources n'ont pas la même nature, ne visent pas la même application et relèvent de différents degrés de formalisation (Aussenac-Gilles, 2004). Et si l'on admet qu'elles visent toutes à représenter des connaissances du domaine, force est de constater que le type de connaissances en jeu varie. Ainsi, un article tel que celui de L'Homme (2008), qui compare trois types de ressources, montre que ces dernières ne contiennent ni la même masse ni le même type d'information.

Par exemple, si les index contiennent essentiellement des listes de termes/concepts, les glossaires leur accolent en plus une courte définition. Les ontologies formalisent quant à elles des termes/concepts, les relations entre ces concepts, des informations sur les propriétés des objets qu'elles contiennent, des instances de concepts, des axiomes pour permettre de raisonner automatiquement sur ces connaissances. Les types de relations conceptuelles enregistrées dans les ressources peuvent également différer : si les relations classiques sont la

³⁷ « Suite à l'utilisation généralisée des outils de bureautique, à l'internationalisation des échanges et au développement d'Internet, la production de documents sous forme électronique s'accélère sans cesse. Or pour produire, diffuser, rechercher, exploiter et traduire ces documents, les outils de gestion de l'information ont besoin de ressources terminologiques. La gamme des produits à base terminologique nécessaires pour répondre à ces besoins s'élargit considérablement » (Bourigault & Jacquemin, 2000 cité par (Bourigault & Aussenac-Gilles, 2003 : 2)).

³⁸ Ces différents besoins et types de ressources sont soutenus par le rapprochement entre terminologie et intelligence artificielle (représenté en France par le groupe TIA, cf. note de bas de page 20). On parle donc de ressources *terminologiques* mais également de ressources *termino-ontologiques* (ou RTO).

relation générique/spécifique et la relation « voir-aussi », certaines ressources prennent en compte les liens de synonymie, de méronymie, etc.

Ces quelques éléments montrent que si l'on considère les langues de spécialité comme des vecteurs de connaissances, on doit prendre conscience de l'hétérogénéité que cette notion implique, ce qu'illustrent Haton *et al.* (1991 : 22) en ingénierie des connaissances :

« le terme de connaissances recouvre les différentes formes du savoir : objet du monde, faits concernant ces objets (la neige est blanche), classifications (par exemple, taxinomies en zoologie ou en géologie), événements (la température du four s'est mise à osciller), règles heuristiques de savoir-faire (si le moteur cale à froid, le gicleur est peut-être bouché), etc. »

Dans la pratique, la délimitation des connaissances en jeu dépend fortement de l'application, de la ressource envisagée, des corpus et d'un ensemble de paramètres externes qui permettent de circonscrire ce que l'on recherche dans les textes. Cette prise en compte vient confirmer le fait que l'extraction de connaissances à partir de textes relève bien d'une construction et d'une interprétation.

Dans notre travail, nous avons choisi de ne pas poser d'*a priori* sur les types de connaissances susceptibles d'évoluer dans le temps. Ce point de vue est justifié par le fait que nous ne disposons que de très peu d'informations sur l'évolution des connaissances en diachronie. Cet aspect est d'autant plus marqué que nous travaillons en diachronie courte, perspective rarement envisagée dans la mesure où elle laisse généralement présumer que peu de changements sont observables. Pour cette raison, nous essayons plutôt de faire « émerger » des textes les connaissances en jeu afin de caractériser ce qu'est l'évolution en diachronie courte et d'en montrer l'hétérogénéité. Notre perspective est donc d'observer comment on accède aux connaissances à partir d'une analyse linguistique et comment on peut les interpréter.

2.1.2.3 Dimension interne : la langue et les textes

La terminologie textuelle place au centre de l'analyse les textes. Comme le note Petit (2001 : 71-72), une telle perspective pose des questions importantes quant au type de textes à prendre en compte et à la définition de ce que sont les textes spécialisés. Des éléments de réponse sont apportés dans le chapitre suivant, mais les textes à prendre en compte pour l'analyse peuvent provisoirement être définis comme :

« ce que les spécialistes disent et écrivent, entendent et lisent pour réaliser la communication de spécialité : entretiens, discours, conférences, débats, réunions ; et lettres, rapports, comptes-rendus, analyses, procès-verbaux, documents de travail, articles, manuels, livres, publications techniques et scientifiques diverses. » (Kocourek, 1991b : 21)

Comme dans le cas de la dimension externe, ce que l'on cherche à observer dans la dimension interne à partir de textes est hétérogène et dépasse le niveau lexical seul. Ainsi que le souligne Kocourek (1982 : 42) par exemple :

« C'est principalement sur la base de ces textes que l'on cherche à saisir le système de la langue de spécialité, à signaler sa délimitation et sa diversification, son fonctionnement, sa spécificité linguistique lexicale, syntaxique, graphique, la formation, signification et structure de sa terminologie, ses perspectives, son appréciation. »

L'analyse linguistique des textes ne se limite donc pas à l'analyse des termes, mais doit prendre en compte l'analyse de la langue de spécialité comme système. L'analyse est donc susceptible de concerner tous les niveaux de description linguistique. Et bien que la terminologie privilégie le niveau lexical, l'accès aux connaissances à partir de textes peut s'appuyer sur des éléments morphologiques, syntaxiques, lexicométriques, etc.

Plus précisément enfin, notons que les textes constituent l'ancrage (ou le support analysable) du discours mais que c'est dans ce dernier que les connaissances sont présentes et discutées. L'importance du discours est centrale en diachronie, dans la mesure où, comme le souligne Benvéniste (1966 : 131), « c'est dans le discours que la langue se forme et se configure. On pourrait dire, en calquant une formule classique : *nihil est in lingua quod non prius fuerit in*

oratione³⁹ ». Cette position est particulièrement bien défendue par la socioterminologie où Gaudin (2003), par exemple, insiste sur le fait que le sens du terme est fabriqué et négocié en discours chaque fois qu'il est utilisé. Pour reprendre les mots de Béjoint (2004 : 385) : « Le discours est [...] le berceau du terme naissant et le lieu où il revient toujours puiser de l'énergie pour de nouvelles aventures ».

La notion de discours peut donc être vue comme l'*interface* entre dimension interne et dimension externe, *interface* à partir de laquelle le lien entre ces deux dimensions peut être reconstruit. Et c'est par l'analyse linguistique des textes⁴⁰, traces des discours, que l'on peut espérer accéder aux connaissances et à leur évolution, au terme d'une analyse linguistique, d'une interprétation de l'analyste, d'une collaboration avec des experts et de la prise en compte d'une application. C'est ce que nous montrons dans la suite de notre recherche.

2.2 Points de vue historique et diachronique

L'objectif de cette section est de montrer la manière dont la distinction entre dimension interne et dimension externe participe à la mise en place de différents points de vue pour l'analyse diachronique, ainsi que la manière dont nous nous positionnons en langue de spécialité. En effet, maintenant que nous avons vu la manière dont les langues de spécialité permettent d'associer interne/externe et de postuler un lien reconstituable entre ces deux dimensions, nous devons préciser la manière dont elles s'articulent en diachronie.

2.2.1 Histoire interne et histoire externe

La distinction saussurienne entre éléments internes et éléments externes de la langue, reprise en diachronie par Marchello-Nizia (1995 : 24), repose sur l'observation du fait que l'évolution des langues peut être appréhendée :

- du point de vue politique et culturel (histoire externe),
- du point de vue du système de la langue (histoire interne).

³⁹ « rien n'est dans la langue qui n'ait d'abord été dans le discours ».

⁴⁰ À condition que le discours soit « situé » (Condamines, 2003a), c'est-à-dire que les textes sélectionnés soient caractérisés par des éléments externes, tels que le genre, les rédacteurs, la langue, etc. Cette question est détaillée au chapitre suivant.

Dans le premier cas, il s'agit de retracer les événements politiques, sociaux ou culturels qui ont conditionné les évolutions de la langue (invasions d'autres peuples, lois et réformes, mais aussi découvertes, inventions, etc.). Dans le second cas, il s'agit d'observer les phénomènes linguistiques du changement en eux-mêmes, sans s'appuyer sur des éléments extralinguistiques de quelque nature que ce soit.

Cette distinction est à l'origine de différents points de vue sur le lien entre linguistique et extralinguistique en diachronie. Ce lien peut être exploité à des degrés divers et fait ainsi émerger plusieurs oppositions, notamment entre *histoire de la langue*, *linguistique historique* et *linguistique diachronique*. En effet, si la distinction entre dimension interne (linguistique) et dimension externe (extralinguistique) est reconnue, l'une ou l'autre peut être privilégiée en fonction des objectifs de recherche. Sur cette base, on peut distinguer notamment :

- la linguistique diachronique (Marchello-Nizia, 1995) où l'on fait abstraction de la dimension externe pour se concentrer sur l'évolution interne du système linguistique ;
- la linguistique historique (*ibidem*) où l'on associe perspectives interne et externe pour retracer l'évolution de la langue : on étudie l'évolution interne du système expliquée en partie par l'histoire externe de ce même système ;
- l'histoire de la langue (Antoine, 1981 ; Brunot, 1967 ; Helgorsky, 1981) où l'on ne s'intéresse qu'à la dimension externe de l'histoire d'une langue (« ses succès et ses revers », Brunot, 1967).

On peut également identifier un autre point de vue sur la distinction interne/externe, selon lequel la langue ne serait plus l'objet des recherches historiques, mais leur « témoin » : il s'agit du point de vue des historiens et historiens des sciences. Ainsi, la langue peut être considérée comme le support de témoignages de l'Histoire (histoire de l'humanité, d'une société, d'une culture, d'une politique, d'une science, etc.). Ce n'est donc plus l'évolution de la langue qui constitue l'objet d'analyse, mais les évolutions historiques extralinguistiques identifiables à partir de l'observation de la langue par les historiens. Ces évolutions sont appréhendées à travers le vocabulaire, les textes, les documents et archives, témoins de l'Histoire.

La centralité de l'étude des textes et de la linguistique en Histoire se retrouve dans de nombreux ouvrages d'historiens (Prost, 1996 ou encore dans le volume édité par Achard, et

al., 1985 ; Conein, 1985 ; Fiala, 1985 qui, en tant que linguistes, décrivent les rapports étroits entre Histoire et linguistique) ou d'historiens des sciences et techniques (Acot, 1999 ; Baudet, 1989). Ces derniers en particuliers insistent non seulement sur l'importance des textes, mais également de la terminologie. Ainsi, par exemple, Baudet (*ibidem* : 56) écrit :

« [...] l'histoire des vocabulaires est utile pour l'historien des sciences et des techniques. Ce que je voudrais vous proposer, c'est l'amorce d'une réflexion qui m'entraînera à vous suggérer que la terminologie est tout à fait centrale dans notre discipline et que nous sommes aujourd'hui réunis pour examiner [...] ce vieux problème [...] du rapport entre langage et pensée. »

Ce qui est intéressant dans cette conception est le fait que les textes en Histoire et histoire des sciences sont considérés non pas comme de simples discours « à écouter », mais plutôt comme contenant des termes et des concepts révélateurs du rapport entre langue et connaissances à une période donnée. Cette perspective est adoptée également par l'historien Prost (1996 : 63,) qui explique :

« Aux discours prononcés le 11 novembre devant les monuments aux morts, l'historien ne demandera pas ce qu'ils disent, qui est bien pauvre et répétitif ; il s'intéressera aux termes utilisés, à leurs réseaux d'oppositions ou de substitutions, et il y retrouvera une mentalité, une représentation de la guerre, de la société, de la nation. »

Ce point de vue permet de mettre en avant l'importance du lexique dans l'accès à la dimension externe pour les historiens et le lien entre terminologie et histoire des connaissances. De ce point de vue, les études sur la créativité lexicale sont les plus remarquables dans leur argumentation sur le parallèle entre dimensions interne et externe. Par exemple, Guilbert (1975 : 57) rapporte la vision de Guiraud sur ce parallèle, qui soutient la convergence des deux dimensions pour expliquer la création lexicale :

« C'est le problème qu'a soulevé P. Guiraud dans son livre *Structures étymologiques du lexique français* : « il y a donc un double déterminisme dans lequel, loin de s'exclure, les causes internes et externes se complètent, le mot étant le résultat d'un impact, d'une pression de l'histoire sur le système. Il apparaît alors que les motivations interne et externe sont l'une et l'autre

nécessaires, mais qu'aucune n'est suffisante : c'est leur convergence qui crée le mot nouveau [...] » (p.6). »

Ces éléments montrent qu'en diachronie, dimensions interne et externe peuvent être associées pour décrire l'évolution. Plusieurs possibilités existent quant à cette association riche et confèrent à la langue différents statuts. Parmi ces derniers, deux sont particulièrement marqués :

- le statut de la langue comme objet de l'évolution lorsque l'on analyse son évolution,
- le statut de la langue comme « témoin » de l'évolution lorsque l'on observe son évolution et postule que celle-ci reflète des évolutions externes.

En langue de spécialité, on peut dans une certaine mesure retrouver ces deux points de vue, qui reflètent des différences importantes quant à la prise en compte de la dimension interne ou externe pour décrire l'évolution de la langue.

Le premier prend essentiellement en compte la dimension interne (linguistique). C'est le cas par exemple de travaux de certains néologues qui se concentrent sur la description des matrices lexicogéniques utilisées pour la formation des termes nouveaux (Sablayrolles, 1996b ; 2000b par exemple), que d'autres nomment matrices terminogéniques pour marquer leur ancrage dans le champ des langues de spécialité (Cormier & Rioux, 1991 ; Portelance, 1987). Dans ce cas, les travaux s'appuient sur les recherches existant en langue générale et il n'est pas rare que les chercheurs soient des linguistes plus que des terminologues (*ibidem*). Mais bien que la description linguistique prime sur la description des connaissances, la dimension externe des langues de spécialité reste toujours présente ne serait-ce que par le découpage initial en domaines.

Le second point de vue pose un parallèle externe/interne et permet d'émettre l'hypothèse d'un lien entre évolution de la langue et évolution des connaissances. C'est le point de vue que nous choisissons dans notre recherche dans la mesure où les possibilités d'association entre langue et connaissances s'avèrent centrales tant dans le cadre des langues de spécialité qu'en diachronie. Nous précisons cette hypothèse dans le paragraphe suivant.

2.2.2 Hypothèse d'un parallèle entre langue et connaissances pour la diachronie

Nous avons vu dans ce chapitre que les langues de spécialité se caractérisent par leur association spécifique entre dimension linguistique et dimension extralinguistique. Cette association est également très prégnante en diachronie et laisse formuler l'hypothèse d'un parallèle entre évolution dans la langue et évolution des connaissances. Ce type de lien a été décrit par exemple par Guilbert (1975 : 15) qui souligne que :

« le changement linguistique répond à la nécessité élémentaire de la connaissance qui épouse les rythmes de l'évolution du monde, à la nécessité de la communication de toute expérience nouvelle. Le langage ne serait pas, s'il ne satisfaisait pas ce besoin vital. »

Aujourd'hui, les évolutions théoriques des langues de spécialité permettent de reprendre cette hypothèse et de l'envisager d'un point de vue méthodologique. Sur la base d'une remarque telle que celle de Cabré (1998 : 141), il est en effet possible de mettre en place une analyse diachronique en terminologie textuelle :

« [...] des textes scientifiques sur le même sujet produits à des moments éloignés dans le temps peuvent présenter des différences intéressantes non seulement au niveau de la conceptualisation de la discipline, mais aussi sur le plan de l'expression. »

À partir de ces éléments, nous adoptons l'hypothèse d'un parallèle entre évolution dans la langue et évolution des connaissances et la développons pour l'adapter à une méthodologie diachronique, linguistique et outillée. Plus précisément, nous posons que l'observation de variations dans les textes permet de repérer des évolutions de connaissances. Au vu des éléments décrits dans ce chapitre, ce parallèle implique plusieurs éléments.

Le premier est que les variations dans la langue sont observables en discours. D'un point de vue méthodologique, l'observation du discours est basée sur l'analyse linguistique de textes du domaine et implique une réflexion importante sur la question de constitution de corpus pour explorer l'évolution.

Deuxièmement, observer dans quelles mesures les textes du domaine reflètent l'évolution des connaissances implique d'élargir le lien entre interne/externe à la prise en compte de la

langue comme système, qui s'incarne dans le discours, et des connaissances sous leurs différents aspects. Ceci signifie que les termes/concepts peuvent servir de base à l'analyse (Chapitre 3) mais que celle-ci ne doit pas rester exclusivement centrée sur ce « noyau ».

Troisièmement, établir un parallèle entre langue et connaissance ne signifie pas qu'un type d'évolution linguistique peut être associé de manière univoque et systématique à une évolution des connaissances. Nous illustrons cet aspect dans notre travail, mais on peut citer à ce stade les remarques de Møller (1998) ou Dury (1997 : 87) qui soulignent qu'il « y a de fréquents « décalages » entre le concept et le terme qu'il dénomme, ce dernier ne reflétant pas toujours les modifications qui s'opèrent au niveau notionnel ». Autrement dit, appliquer l'hypothèse d'un parallèle entre langue et connaissances en diachronie implique de ne pas supposer de liens prédéfinis entre ces deux dimensions, mais plutôt de chercher à comprendre quels types de liens on peut reconstruire et la manière dont on peut les reconstruire.

Enfin, ce type d'hypothèse permet d'envisager la mise en place d'une démarche reproductible pour une description de l'évolution de n'importe quel domaine et sur n'importe quel intervalle temporel, dont la diachronie courte. Ce dernier point n'est pas trivial dans la mesure où l'évolution est bien susceptible d'intervenir sur de très courts intervalles et les experts n'ont que peu de recul et de conscience sur ces évolutions rapides. De ce point de vue, la mise en place de projets spatiaux est un excellent exemple.

Dans la suite de notre travail, nous nous concentrons sur la mise en place de la démarche d'exploration en corpus de l'évolution des connaissances pour mettre en œuvre et valider cette hypothèse. Le prochain chapitre a pour objectif de décrire la constitution des corpus textuels pour une analyse diachronique des langues de spécialité, ainsi que la démarche de linguistique outillée dans laquelle nous nous inscrivons.

Chapitre 3 Méthodologie pour une analyse de l'évolution en corpus

SOMMAIRE DU CHAPITRE

3.1	PRINCIPES MÉTHODOLOGIQUES GÉNÉRAUX POUR L'ANALYSE DE L'ÉVOLUTION	62
3.1.1	<i>Démarche comparative : Périodisation et organisation temporelle des corpus</i>	62
3.1.1.1	Synchronie dynamique et diachronie	62
3.1.1.2	Organisation temporelle des corpus : Choix des périodes et intervalles.....	63
3.1.1.3	Trois types de corpus : Corpus synchroniques, corpus de suivi et corpus comparables	67
3.1.2	<i>Démarche outillée : Linguistique « à l'instrument » pour observer le changement</i>	70
3.1.2.1	Linguistique outillée : de nouveaux moyens et de nouveaux besoins.....	70
3.1.2.2	Outillage et indices	73
3.2	CHOIX MÉTHODOLOGIQUES POUR NOTRE ÉTUDE.....	77
3.2.1	<i>Corpus pour le domaine spatial</i>	77
3.2.1.1	Choix des sources	77
3.2.1.2	Domaine d'activité et domaine de connaissance	79
3.2.1.3	Contraintes de constitution des corpus diachroniques.....	82
3.2.2	<i>Description des corpus sélectionnés</i>	84
3.2.2.1	Corpus TTVS	84
3.2.2.2	Corpus DORIS.....	88
3.2.3	<i>Choix des outils et traitement des corpus</i>	91
3.2.3.1	Choix des outils.....	91
3.2.3.2	Extraction et préparation des données.....	97

Tel que décrit dans les chapitres précédents, nous inscrivons notre démarche en terminologie textuelle. Cette position place au centre de notre approche la question de la constitution et de l'exploration de corpus à l'aide d'outils. Ceci implique des choix méthodologiques propres à la diachronie pour garantir la représentativité des textes par rapport au repérage de l'évolution.

Dans ce chapitre, nous présentons ces choix en deux temps. Dans la première section, nous discutons des deux principes méthodologiques essentiels qui sous-tendent notre démarche : les choix liés à la périodisation des corpus pour la mise en place d'une démarche comparative et les choix liés à la description linguistique outillée pour observer l'évolution. Dans la seconde section, nous illustrons les principes de constitution de corpus à travers l'exemple des textes sur lesquels se fonde notre étude.

3.1 Principes méthodologiques généraux pour l'analyse de l'évolution

La constitution de corpus pour l'analyse linguistique et le traitement informatique des textes suscite de nombreuses questions, notamment de représentativité, de taille, etc. (Péry-Woodley, 1995). Ces difficultés sont aujourd'hui connues et assez bien balisées, que ce soit en langue générale ou en langue de spécialité, comme l'atteste l'existence de plusieurs ouvrages sur ces questions (par exemple ceux de Biber, 1993 ; Habert, *et al.*, 1997 ; McEnery & Wilson, 2001 pour la constitution de corpus en langue générale ou Bowker & Pearson, 2002 ; Pearson, 1998 pour les langues de spécialité).

Les corpus diachroniques n'échappent pas à ces contraintes, mais ils doivent prendre en compte d'autres critères propres à l'analyse diachronique et rarement décrits dans ces ouvrages. Dans cette section, nous proposons de nous concentrer essentiellement sur ces questions spécifiques, liées à l'aspect comparatif et à la question de l'outillage pour l'exploration des corpus diachroniques.

3.1.1 Démarche comparative : Périodisation et organisation temporelle des corpus

Observer l'évolution en corpus signifie de chercher à repérer différents états de langues et de connaissances dans les textes. Pour ce faire, la constitution des corpus doit mettre en œuvre des choix centraux concernant la périodisation des textes pour observer le changement. Pour aborder cette question, un retour sur les notions de *synchronie* et *diachronie* s'impose.

3.1.1.1 Synchronie dynamique et diachronie

Une première remarque concerne la distinction entre synchronie et diachronie, distinction essentiellement établie sur l'opposition entre état de langue et phase d'évolution. Cependant, Saussure (1995 : 142) souligne à propos de la synchronie qu'« en pratique, un état de langue n'est pas un point, mais un espace de temps plus ou moins long pendant lequel la somme des modifications survenue est minime. Cela peut être dix ans, une génération, un siècle ». La définition d'un état de langue (d'une synchronie) repose donc sur l'importance ou non des changements observables. Synchronie et diachronie sont de fait avant tout des abstractions méthodologiques nécessaires pour mener des études linguistiques, l'une pour observer un état de langue et faire abstraction des changements (« minimales »), l'autre pour comparer

différents états successifs d'une langue et en décrire l'évolution. De fait, la diachronie est la perspective privilégiée pour observer le changement.

Cependant, il est clair qu'aborder la description de la langue du point de vue synchronique revient à mettre entre parenthèses le changement linguistique et faire de la langue un objet figé. Pour ne pas ignorer les modifications qui affectent incontestablement la langue à tout moment de son histoire, les fonctionnalistes entre autres ont proposé le concept de *synchronie dynamique* (Martinet, 1990 ; Houdebine, 1985 : 7). La synchronie dynamique repose sur « la coexistence éventuelle d'usages différents d'une entité dans le discours, à un moment donné » (Feuillard, 2007 : 5) et prend son origine dans les études du changement phonétique (Feuillard & Krier, 2000 par exemple). Dans ce cadre, l'objectif est d'interroger différents « chronolèctes » à travers la mise en place de sondages de locuteurs de générations différentes. Bien que les fonctionnalistes s'intéressent principalement aux changements phonétiques, le principe de *synchronie dynamique* se retrouve dans d'autres travaux, et notamment ceux portant sur la néologie. Par exemple Gaudin (1993 : 163) souligne :

« Pour comprendre le mouvement de la néologie, il ne faut pas la concevoir dans une optique strictement synchronique mais la situer, en dépassant l'opposition classique synchronie/diachronie, dans le cadre d'une synchronie dynamique. C'est là un aspect majeur de la terminologie, dans la mesure où la synchronie qu'elle étudie est toujours prise dans le mouvement qui est le « lieu de rencontre entre l'innovation et l'archaïsme » (Guilbert, 1973b) ».

Le concept de *synchronie dynamique* se définit à travers la coexistence en synchronie d'usages anciens et nouveaux permettant de saisir l'évolution de la langue. Il existe donc deux approches pour observer l'évolution dans la langue : une approche diachronique, où l'on compare différents états de langue, et une approche synchronique dynamique, où l'on compare différents usages dans un même état de langue. Dans les deux cas, les choix liés à l'organisation temporelle des corpus sont centraux.

3.1.1.2 Organisation temporelle des corpus : Choix des périodes et intervalles

L'organisation temporelle des corpus pour observer l'évolution est établie en fonction de deux critères : le premier concerne le choix de la *période* à observer, c'est-à-dire l'époque pour

laquelle on cherche à décrire l'évolution, et le second concerne l'*intervalle* de temps à considérer, c'est-à-dire le laps de temps pendant lequel les changements sont susceptibles de se produire.

Helgorsky (1981 : 124) distingue deux approches pour établir ces choix. La première repose sur une chronologie externe : les corpus à observer sont définis en fonction d'événements politiques, historiques, sociaux dont on fait l'hypothèse qu'ils sont susceptibles de se refléter dans la langue. Par exemple, Drouin *et al.* (2006) ont choisi de comparer des corpus de textes parus avant et après le 11 septembre 2001 pour observer la néologie dans la terminologie du terrorisme. Ahmad et Musacchio (2004), quant à eux, ont défini trois grandes périodes de l'évolution de la terminologie de la physique nucléaire auxquelles ils associent des types de textes dans lesquels sont diffusées ces connaissances⁴¹. Cette perspective se rapproche d'une autre de leurs propositions (Ahmad, 1996) encore une fois dans le domaine de la physique, et qu'adopte Chateau (2007) pour le domaine de la géologie : le découpage temporel des corpus est établi en fonction de grandes périodes identifiées dans les domaines comme étant des paradigmes scientifiques (Kuhn, 1999). Les travaux de Chateau cherchent d'ailleurs spécifiquement à explorer dans quelles mesures les corpus et la langue reflètent les ruptures entre paradigmes et soutiennent ainsi la théorie de Kuhn.

La seconde approche identifiée par Helgorsky repose sur une chronologie interne. Selon cette approche, il s'agit d'établir un découpage temporel à partir de différents états de langues prédéfinis tels que l'Ancien Français, le Moyen Français, le Français Classique ou le Français Contemporain dont on sait *a priori* qu'ils présentent des différences linguistiques fondamentales. C'est l'optique choisie par exemple par Marchello-Nizia (1995 ; 2006), malgré le fait que la définition de ces différents états de langue n'aille pas toujours de soi (Prévost, 2008).

Au vu de ces deux critères et de ces quelques exemples d'études, la chronologie externe apparaît comme la perspective privilégiée pour aborder la question de l'évolution en langue

⁴¹ « The timeline is divided into three major zones : discovery, adaptation and utility. The discovery phase is characterized by writing in the formal genre, that of learned papers, the adaptation phase is where the knowledge has been researched to a good degree and has been adapted for teaching and learning about the discovery and, in addition to the learned text, instructional text is also created. The third utility phase is where the discovered knowledge is routinely used and this knowledge is reported in the less formal genre of newspaper reportage and popular science literature » (Ahmad et Musacchio, 2004 : 1567).

de spécialité où l'objectif est généralement lié à la description des connaissances. Dans notre approche, nous illustrons cette démarche puisque pour construire nos données, nous partons de l'identification de projets spatiaux ou de périodes dans le domaine spatial dont on fait l'hypothèse qu'ils reflètent des évolutions de connaissances dans les textes.

Le choix de l'intervalle temporel pendant lequel explorer l'évolution, quant à lui, est une question difficile, en particulier dans la mesure où la vitesse de l'évolution d'une langue est peu prévisible :

« une langue changera à peine pendant un long intervalle pour subir ensuite des transformations considérables en quelques années. De deux langues coexistant dans une même période, l'une peut évoluer beaucoup et l'autre presque pas [...]. » (Saussure (De), 1995 : 142)

Ce que l'on sait en tout cas, c'est que les niveaux de description linguistique n'évoluent pas tous au même rythme. Ainsi, on remarque souvent que le lexique évolue plus vite que la syntaxe (Mair & Leech, 2006) ou encore que les changements phonétiques sont généralement longs. Néanmoins, à l'intérieur de chaque niveau de description, les changements n'interviennent pas tous au même rythme (Helgorsky, 1981) et certains changements phonétiques peuvent ainsi être plus rapides que d'autres. Sur la base de ces observations et en fonction des objectifs de la description, on pourra prévoir des intervalles plus ou moins longs : un néologue pourra envisager d'observer des intervalles courts (quelques années), alors qu'une étude sur l'évolution phonologique demandera plutôt des intervalles de plusieurs décennies, voire siècles.

Du point de vue externe on ne dispose que de peu d'éléments sur la rapidité de l'évolution des connaissances qui permettent de choisir aisément un intervalle. Une solution possible est de choisir l'intervalle de temps à observer en fonction de la période définie pour l'observation. Par exemple, Drouin *et al.* (2006) qui cherchent à observer les évolutions terminologiques du domaine du terrorisme avant et après les attentats du 11 septembre, envisagent un intervalle court de 10 ans (1995-2005). Leur corpus est découpé en deux sous-corpus : « avant 2001 » et « après 2001 » pour pouvoir associer les changements

terminologiques à l'activité terroriste de 2001. Dans ces conditions, l'intervalle peut être très court mais très pertinent pour un objectif donné.

Cet aspect doit être mis en lien avec un certain biais apparu dans les recherches diachroniques. En effet, puisque dans la perspective saussurienne la synchronie et la diachronie sont corrélées au degré de changement⁴², les études diachroniques prennent essentiellement en compte des intervalles très longs, généralement de plusieurs siècles. Puisqu'il est difficile de savoir *a priori*, si une langue a subi *assez* de changements pour être observée en diachronie, une solution est donc de prendre systématiquement en compte des intervalles de temps très longs, au cours desquels on est certain de repérer des changements.

Ce point de vue, largement adopté par les diachroniciens, est tout à fait justifiable mais a beaucoup limité les études consacrées à l'évolution, en particulier en « diachronie courte ». En effet, les études en diachronie courte (ou encore *brachychrony* (Mair, 1997), §1.1.2.3), reposent sur de courts intervalles temporels et ont souvent du mal à trouver une légitimité parmi les historiens et diachroniciens de la langue. Ces derniers considèrent en effet que les changements observés sont trop minimes pour mériter une description diachronique.

Pourtant, si l'on adopte un point de vue externe et que l'on s'attache à définir des intervalles en fonction d'événements politiques, scientifiques ou culturels, la diachronie courte s'impose. Dans le cadre des langues de spécialité en particulier, les progrès scientifiques et techniques sont souvent très rapides et entraînent une croissance suffisamment importante dans les domaines pour que les changements apparaissent à très court terme dans leur terminologie. Ainsi, par exemple, un événement comme les attentats du 11 septembre a pu entraîner en quelques années de nombreuses créations terminologiques dans le domaine du terrorisme (Drouin *et al.*, *op.cit.*). Le domaine de l'écologie a montré lui aussi de nombreux changements terminologiques rapides ces dernières années (notamment au niveau de la vulgarisation et de la terminologisation (Dury, 2008b ; Dury & Lervad, 2007)). De fait, dans une perspective telle que la nôtre où la constitution des corpus dépend majoritairement de critères externes, la diachronie courte est la perspective qui s'impose. De ce point de vue, les langues de spécialité offrent donc un champ de recherche original pour prendre en compte des intervalles

⁴² Rappelons que pour Saussure, si l'on recense peu de changements l'étude doit être synchronique, mais si les changements sont importants, l'étude doit être diachronique.

d'observation de tous ordres et proposer des réflexions nouvelles sur l'évolution, des connaissances et de la langue.

En fonction du choix des périodes et intervalles définis, trois types de corpus peuvent être envisagés, corpus que nous définissons dans le paragraphe suivant.

3.1.1.3 Trois types de corpus : Corpus synchroniques, corpus de suivi et corpus comparables

Pour observer le changement, on peut d'abord se baser sur des corpus synchroniques. En effet, une fois la période d'observation de l'évolution sélectionnée, on peut adopter une démarche synchronique dynamique et chercher à observer différents usages et états de connaissances coexistants dans une synchronie. Nous en donnerons un exemple au Chapitre 5 notamment. Néanmoins pour observer l'évolution, l'approche généralement privilégiée repose sur la comparaison de différents états de langue et de connaissances à partir de corpus diachroniques. Bien qu'une approche diachronique ne se réduise pas à décrire des synchronies successives (Prévost, 2008) d'un point de vue méthodologique, observer l'évolution s'inscrit essentiellement dans une démarche comparative. Ceci signifie que l'on fonde l'observation de l'évolution sur la comparaison d'une succession de synchronies, ce que certains regrettent, tels que Fiala (1985 : 76 cité dans De Surmont, 2005 : 130-131) :

« Dans la représentation de l'histoire qu'ont les linguistes, celle-ci est souvent assimilée à l'idée qu'ils se font de la diachronie, c'est-à-dire une succession, une somme de synchronies, alors qu'elle est tout autre chose, quelque chose de beaucoup plus riche qui implique la dynamique du changement, c'est-à-dire une représentation plus continue du temps. »

La comparaison de différents états de langue est cependant inévitable. Néanmoins, il est possible de réduire l'intervalle de temps considéré entre les différentes coupes synchroniques pour tenter de saisir au plus près le dynamisme de la langue. En fonction de cet intervalle, deux types de corpus peuvent être envisagés : les corpus de suivi et les corpus comparables.

Les corpus de suivi⁴³ regroupent des « flux continus de textes permettant l'analyse chronologique, année par année par exemple, de données langagières » (Habert, *et al.*, 1997 : 125). En d'autres termes, il s'agit de corpus ouverts alimentés périodiquement en textes fraîchement publiés. D'un point de vue diachronique, l'intérêt de ce type de corpus est qu'il permet d'adopter une représentation plus continue du changement⁴⁴. Les corpus de suivi sont à l'heure actuelle généralement constitués d'articles de journaux dont on suppose souvent qu'ils reflètent « [...] l'histoire au jour le jour » et que « l'étude de la langue à travers les journaux, c'est l'histoire de la langue au jour le jour » (Antoine, 1981 : 155). Bien que la question de la représentativité des corpus journalistiques mérite d'être posée (Bourigault & Galy, 2005), ces derniers sont aujourd'hui parmi les seuls textes permettant d'obtenir des flux continus et réguliers pour assurer un corpus de suivi cohérent.

À titre d'exemple, dès le début des années 90, Renouf (1993) travaille sur un corpus constitué des différentes éditions du *Times* de novembre 1990 à septembre 1991 pour observer mois après mois la néologie, la productivité de certains mécanismes morphologiques, les nouvelles collocations, etc. Ce type de corpus a été envisagé dans le cadre du projet AVIATOR (Blackwell, 1993 ; Collier, 1993 ; Renouf, 1993), dont le but était de créer un système capable de saisir l'évolution de la langue au cours du temps pour l'anglais. Un autre exemple de projet mettant à contribution un corpus de suivi est celui du système Glossanet, mis en place au LADL⁴⁵ en 1999 (Fairon, 2000) et qui « permet de réaliser des opérations de veille, mais aussi faire de l'extraction d'information et de rechercher des structures lexicales ou syntaxiques » (Fairon & Courtois, 2000). Ces corpus sont donc très utiles aux études néologiques pour mettre en place de véritables contextes de « veille néologique » (Cartier & Sablayrolles, 2008) et ainsi améliorer et assister la mise à jour de dictionnaires. De plus, avec l'émergence de l'utilisation du Web comme corpus, l'exploitation de corpus de suivi est de plus en plus souvent envisagée⁴⁶.

⁴³ Appelés aussi « monitor corpus » (Sinclair, 1991 : 4), « corpus baromètres » (Habert, *et al.*, 1997 : 145), « corpus dynamiques » (Renouf, 1994) ou « corpus ouverts » (Bowker & Pearson, 2002 : 12-13).

⁴⁴ Bien que rien n'empêche de considérer des intervalles de temps plus longs. Les termes « monitor corpus » et « corpus ouverts » sont en effet parfois utilisés pour renvoyer à des corpus que l'on complète régulièrement, mais qui ne prennent pas en compte des intervalles courts ou réguliers.

⁴⁵ Laboratoire d'Automatique Documentaire et Linguistique.

⁴⁶ On notera à titre d'exemple l'émergence de prototypes tels que le Chronologue de Véronis (<http://aixtal.blogspot.com/2005/12/outil-le-chronologue.html>, consulté le 12 décembre 2008).

Enfin, comme le notent Habert *et al.* (1997 : 145), dans la mesure où les corpus de suivi sont récents, ils ne renseignent encore que sur la courte durée (deux décennies au plus), mais au fur et à mesure que le temps passe, leurs données s'accroissent et laissent envisager de mener de nouvelles descriptions de l'évolution des langues.

Cependant aujourd'hui, les corpus de suivi sont essentiellement exploités dans le cadre de l'évolution en langue générale et posent encore de nombreuses difficultés méthodologiques et théoriques pour observer le changement (Habert, *et al.*, 1997). De plus, leur constitution reste assez difficile et limitée dans la mesure où ils nécessitent la prise en compte de flux de textes encore peu répandus, en particulier pour les langues de spécialité.

Le deuxième type de corpus diachroniques qui permet d'observer l'évolution, beaucoup plus répandu, est celui des corpus comparables⁴⁷. Les corpus comparables sont des sélections de textes similaires dans différentes langues ou dans plusieurs variétés d'une même langue (Sinclair, 1991 : 12). Ce type de corpus est généralement mis en place dans le cadre de la traduction, de la comparaison de langues (L'Homme, 2004 : 126 ; Bowker & Pearson, 2002 : 93) ou d'extraction de lexiques bilingues (Dejean & Gaussier, 2002 par exemple). Dans le cas d'une description diachronique, il s'agit de textes de genre, domaine, langue similaires, mais produits à différentes périodes :

« The ideal situation for comparing corpora as a basis for studying language change would appear to be to use two corpora constructed on parallel principles at two different points in time. Assuming that any variation identified can be reasonably attributed to language change over time, rather than to say, topic differences or stylistic differences between the corpora, then comparing two similarly constructed corpora seems to offer a relatively straightforward method of identifying at least lexical changes which have occurred during the period between the two collection dates. » (Holmes, 1994 : 28)

Il s'agit généralement de corpus fermés pour la comparaison de plusieurs états de langue à des périodes différentes. Les intervalles considérés sont plus longs que dans le cas des corpus de suivi puisque l'on ne cherche pas à accéder à un flux rapide et quasi-continu de l'évolution de la langue mais à une comparaison de différents états « clos » pour décrire le changement.

⁴⁷ Ou corpus chronologiques (Calberg-Challot, *et al.*, 2006).

On supprime donc la dimension de veille propre aux corpus de suivi. Des exemples de constitution de corpus diachroniques comparables sont donnés notamment par Kytö (2000), Meurman-Solin (1995) ou Holmes (1994) pour la langue générale et Dury (2004) pour les langues de spécialité. Ces différents auteurs insistent sur les difficultés propres à la constitution de ce genre de corpus dont le choix des périodes et intervalles, celui des genres des textes, celui de la langue, de la taille du corpus et la disponibilité des sources, questions que nous détaillons pour notre étude (§3.2.1.3, p.82).

Dans tous les cas, et pour suivre le principe « not chance but choice » d'Engwall (1994), le choix du corpus doit faire l'objet d'un soin particulier pour s'assurer de sa représentativité et de la possibilité de corrélérer les résultats obtenus à l'évolution de la langue/des connaissances. La construction d'un corpus doit en effet faire l'objet d'une démarche raisonnée et non pas d'un « regroupement opportuniste » (Habert, *et al.*, 1998 : 35).

3.1.2 Démarche outillée : Linguistique « à l'instrument » pour observer le changement

La démarche proposée dans notre approche se fonde sur l'utilisation d'outils pour l'analyse et le traitement des corpus. Pour mieux comprendre ce choix, nous définissons ce que l'outillage implique, notamment dans une perspective diachronique telle que la nôtre.

3.1.2.1 Linguistique outillée : de nouveaux moyens et de nouveaux besoins

Comme l'écrit Ho-Dac (2007 : 141)

« une étude linguistique en corpus décrit des fonctionnements linguistiques en se basant sur une interprétation de faits réellement observés. Cette approche repose sur l'idée de partir à la découverte, d'aller explorer des réalités langagières pour y découvrir des régularités linguistiques. »

C'est précisément dans ce cadre de découverte que nous concevons notre approche diachronique, à travers le repérage et l'observation de variations dans les textes que l'on peut associer à des évolutions de connaissances.

L'« exploration » de fonctionnements linguistiques en corpus va systématiquement de pair avec l'utilisation d'outils informatiques, donnant naissance à ce qu'Habert (2005 : 1) nomme une *linguistique « à l'instrument »*. L'auteur souligne en effet qu'à côté d'une linguistique « sans instrument »⁴⁸, s'impose clairement une linguistique qui utilise des outils⁴⁹ (concordanciers, étiqueteurs, etc.) et des ressources (bases textuelles, corpus de textes, dictionnaires, etc.) pour construire des analyses. Dans cette perspective, se développe une coopération riche entre linguistique et Traitement Automatique des Langues (TAL), les travaux en linguistique ayant « beaucoup à gagner à intégrer les outils fournis par le TAL et par l'informatique en général » (Habert, 2004, suivi notamment par Gasiglia, 2004 ; Leroy, 2004 ou Méla 2004). C'est dans le champ de cette linguistique outillée, qui propose une collaboration entre outils de TAL et corpus, que nous inscrivons la méthode linguistique proposée dans notre étude⁵⁰.

Bien que l'outillage ne soit pas « obligatoire » pour l'analyse des textes (Condamines, 2003 : 79), il est indéniable que le recours à des outils change le regard du linguiste sur les données et enrichit l'analyse sur plusieurs points. Tout d'abord, les outils permettent de traiter de plus grandes masses de textes, plus rapidement. Mais surtout, comme le note Condamines (*ibidem*), les outils assistent le travail du linguiste et offrent des angles d'observation particuliers sur certains phénomènes langagiers en « délinéarisant » l'analyse du texte. Ils

⁴⁸ Pour illustrer cette linguistique « sans instrument », dans une communication présentée à l'Inalco en 2006 (www.limsi.fr/Individu/habert/Paroles/BHabertInalco06InstrumentsEnLinguistique-psnup-l4.pdf, consulté le 1^{er} mai 2009), Habert cite Aurox (1998 : 170) : « L'une des particularités des sciences du langage... c'est que le langage est sans médiation à disposition du lecteur : je puis produire, à volonté, des phrases, les tronquer, y introduire tel élément que je choisis, etc. Il se pourrait que ce soit le seul exemple d'une manipulation sans instrument, du moins le seul qui se soit maintenu dans un état développé d'une discipline scientifique. »

⁴⁹ Nous n'adopterons pas ici la distinction proposée par Habert entre *outil*, *instrument* et *utilitaire* (2005 : 2-3) et simplifions la vision de l'auteur en utilisant « linguistique instrumentée » et « linguistique outillée » comme équivalents.

⁵⁰ Notons qu'il peut être difficile de trouver sa place dans le champ de la linguistique de corpus. En effet, comme le note entre autres Mayaffre (2005) « tout le monde utilise le terme *corpus* » et il peut être délicat de se positionner. Dans notre cas, ce phénomène est accentué par le fait que les réflexions les plus poussées et solides dans ce champ sont celle de la *linguistique de corpus* telle qu'elle est définie depuis quelques années dans le monde anglo-saxon (Biber, 1993 ; McNery & Wilson, 2001 ; Tognini-Bonelli, 2001) où sont mis en avant les principes de quantification, de généralisation et ce qui implique la prise en compte de grosses masses de textes (Péry-Woodley, 2005). Ces aspects sont difficiles à appréhender, en particulier en langue de spécialité et en diachronie. Notre démarche repose en effet sur de petits corpus (comme c'est généralement le cas en langue de spécialité et en diachronie (Bowker & Pearson, 2002 : 45)), mais se veut assez représentative, bien que la quantification de données soit limitée par la taille des données construites. Il s'agit là certainement d'une piste de réflexion à développer.

permettent ainsi de rapprocher des éléments éloignés dans les textes mais qui partagent des fonctionnements linguistiques semblables (forme, distribution, etc.). Cette démarche permet de mettre en évidence des phénomènes que l'on n'aurait pu observer à l'œil nu. C'est ce que montre Habert lorsqu'il parle des outils comme des « créateurs de phénomènes »⁵¹, reformulé plus globalement par Valette (2008 : 11) qui place la coopération entre possibilités d'instrumentation et analyse linguistique dans un « cercle vertueux » :

« L'instrumentation, constitutive de la linguistique de corpus, donne lieu à ce que nous pourrions appeler son « cercle vertueux ». Les grandes masses de données textuelles ou documentaires nécessitent, pour être analysées et décrites, des dispositifs expérimentaux et des instruments *ad hoc*. Cette instrumentation permet de construire de nouveaux observables qui seraient demeurés invisibles autrement. »

Nous l'avons vu dans les chapitres précédents, ce cercle vertueux se traduit par l'émergence de nouvelles propositions théoriques en terminologie, ainsi que par l'apparition de nouveaux besoins pour lesquels l'analyse de l'évolution est centrale. Ces besoins sont d'une part descriptifs, dans la mesure où l'on ne dispose que de très peu de descriptions diachroniques en langue de spécialité, et d'autre part applicatifs, puisque l'on peut envisager d'automatiser (au moins en partie) des tâches telles que la mise à jour de ressources, la veille, etc. La description de l'évolution en diachronie courte naît également en grande part de ce cercle vertueux dans la mesure où c'est la coopération entre connaissances linguistiques et outillage qui permet d'envisager aujourd'hui d'observer les phénomènes d'évolution très fins susceptibles de se manifester sur de courts intervalles.

De manière générale, la collaboration entre la prise en compte de données textuelles et analyses outillées est centrale dans une approche diachronique en langue de spécialité et ce, à double titre. Tout d'abord, dans une perspective historique et diachronique, le recours aux données attestées est primordial pour observer le changement. En effet, sur les états de langue anciens l'analyste ne possède pas de *compétence de production*, mais acquiert plutôt une *compétence de reconnaissance* à force de travailler sur les langues anciennes (Marchello-Nizia,

⁵¹ Dans une communication présentée à l'Inalco en 2006 (www.limsi.fr/Individu/habert/Paroles/BHabertInalco06InstrumentsEnLinguistique-psnup-l4.pdf, consulté le 1^{er} mai 2009).

1995 citée par Lebel, 2003 : 56-57). Dans ce cadre, il n'est donc pas possible d'envisager une approche intuitive sur la langue, comme on peut le faire en linguistique synchronique. En langue de spécialité, le même type d'observation peut être fait à propos des connaissances du domaine étudié, dans lequel l'analyste n'a pas ou n'a que peu de compétence, que ce soit en synchronie ou, *a fortiori*, en diachronie. De fait, en langue de spécialité et en diachronie, les outils permettent effectivement de mettre en évidence des phénomènes linguistiques centraux que les approches habituelles ne permettent pas de mettre au jour, soit parce que le locuteur n'a pas de compétence sur l'état de langue/connaissances observé, soit parce que les phénomènes observables sont trop fins et très peu décrits⁵².

Mais une fois ce constat fait, il reste à savoir comment mettre en œuvre une approche linguistique outillée pour la diachronie en langue de spécialité.

3.1.2.2 Outillage et indices

Pour reprendre les termes de Ho-Dac, citée plus haut, l'approche sur corpus repose sur l'idée d'explorer les textes pour mettre au jour des régularités linguistiques. En lien avec notre hypothèse, lorsque ces régularités sont associables à des évolutions de connaissances, elles sont appelées *indices*.

Pour découvrir ces régularités, deux questions se posent : quels types de régularités doit-on mettre au jour pour observer l'évolution? Avec quelle démarche?

Pour répondre à la première question, nous nous fondons sur deux observations. La première a été formulée par Lass (1980 : 95) qui remarque que « it is impossible to observe the exact moment of change, particularly in speech, but [...] the inference of change is feasible ». La seconde, de Renouf (2002 : 29), liste différentes expressions du changement linguistique : « Change manifests itself in new coinage, in the spread of a feature, in patterns and degrees of productivity, in the gradual assimilation of a new feature into the conventional lexicon, or its eventual departure ». Ces deux remarques permettent de mettre en avant :

⁵² Marchello-Nizia (2004 : 58) remarque d'ailleurs que « peu de décennies après les débuts du traitement automatique des données linguistiques, et une quinzaine d'années après qu'il est devenu relativement facile d'accès aux linguistes, il n'est plus de linguistique historique qui se passe de ce type de pratique, soit à travers une manipulation directe des corpus, soit par le recours à des données quantifiées qui en sont issues. »

1. que l'on n'observe pas le changement en cours, mais que l'on peut observer des traces du changement, traces à interpréter ensuite pour reconstruire l'évolution ;
2. que les régularités doivent être définies en fonction d'hypothèses sur le fonctionnement linguistique, que seule une compétence de linguiste permet de mettre au jour.

Ces deux remarques permettent de rappeler l'importance de l'interprétation dans l'analyse et le fait que les régularités observables en corpus ne le sont que sous réserve de pouvoir *construire* une interprétation à partir des données. Dans la mesure où notre approche est une approche linguistique en langue de spécialité, l'interprétation doit revenir d'une part au linguiste/terminologue qui mène l'analyse, d'autre part à des experts du domaine qui acceptent de collaborer avec lui, ce que nous détaillons dans la suite de notre recherche.

Mais, sur la base de ces observations, il faut mettre en avant le fait que les régularités que l'on cherche à observer dans notre approche s'établissent sur deux niveaux. Premièrement, ces régularités interviennent en diachronie. Dans un second temps, il s'agit de voir si elles peuvent être ou non associées régulièrement à des évolutions de connaissances dans le domaine. Dans le premier cas, c'est la compétence du linguiste/terminologue essentiellement qui permet de statuer sur la régularité des phénomènes linguistiques. Dans le second, c'est l'analyse conjointe du linguiste/terminologue et des experts qui permet de statuer sur la régularité du lien entre fonctionnement linguistique et évolution des connaissances, c'est-à-dire sur le statut d'indice des régularités de fonctionnement mises au jour.

La démarche que nous proposons pour repérer ces types de régularités se situe dans un aller-retour entre les corpus et des hypothèses sur le changement. Plus précisément, nous partons de notre connaissance des fonctionnements linguistiques pour formuler une hypothèse que nous observons en corpus. En fonction des résultats obtenus, et en accord avec le jugement des experts, cette hypothèse est ajustée et complétée jusqu'à la mise au jour de régularités satisfaisantes que l'on pourra alors qualifier d'*indices linguistiques* pour repérer l'évolution des connaissances.

Dans notre étude, comme nous le verrons dans la partie suivante, la première hypothèse posée concerne l'observation des variations de fréquence des termes comme indice

d'évolution : sur la base des travaux en néologie, et sur la base du fait que la fréquence est le « moyen d'entrée » dans les textes quasi-systématiquement mis en œuvre en corpus, nous posons l'hypothèse que des variations de fréquence des termes peuvent être associées à des évolutions de connaissances. Comme détaillé au Chapitre 4, une description de l'indice de fréquence a été mise en œuvre, description qui nous a permis de faire émerger d'autres indices, à leur tour observés en corpus. Cet aller-retour entre les données et les hypothèses nous a permis de mettre au jour quatre indices que nous détaillons dans la partie suivante :

1. les empreintes de fréquence, c'est-à-dire certaines variations de fréquence des termes dans le temps,
2. les contextes riches en connaissances évolutives, c'est-à-dire des portions de textes qui contiennent des informations pertinentes à extraire concernant l'évolution des connaissances,
3. les variantes terminologiques, c'est-à-dire la coexistence ou la succession de dénominations équivalentes pour un même concept,
4. les dépendances syntaxiques, c'est-à-dire des changements dans la cooccurrence syntaxique de certains termes dans le temps.

Du point de vue de l'outillage informatique pour mettre en œuvre ce type de démarche, il faut souligner qu'aujourd'hui, il n'existe pas, à notre connaissance, d'outils de traitement de corpus dédiés à la perspective diachronique⁵³. Les progrès les plus notables en TAL à ce jour et qui se répercutent en diachronie résident essentiellement dans la constitution et la gestion de corpus de textes de plus en plus volumineux⁵⁴. Les analyses diachroniques doivent donc reposer sur des outils classiques à disposition de la linguistique « synchronique outillée ». Les recherches diachroniques instrumentées procèdent ainsi en grande partie à la mise en place de chaînes de traitement pour l'analyse à l'aide d'outils et de logiciels « génériques » en linguistique de corpus tels que des concordanciers, bien que ceux-ci ne puissent pas toujours

⁵³ À l'exception des systèmes de repérage des néologismes ou de « veille néologique ».

⁵⁴ Ces progrès sont particulièrement importants pour les états anciens de langue tels que le Français Médiéval grâce aux nouvelles possibilités de numérisation et d'océrisation des textes anciens (Heiden & Lavrentiev, 2004 ; Marchello-Nizia, 2002 ; Prévost, 2005). L'étiquetage morpho-syntaxique est également de plus en plus performant pour annoter les états anciens de langue (Heiden & Prévost, 2002).

répondre aux contraintes et besoins comparatifs propres à une démarche comparative⁵⁵. Mais comme le soulignent Habert et Fuchs (2004 : 1) :

« pour l'utilisateur (qu'il soit linguiste ou bien spécialiste d'une autre discipline le conduisant à travailler sur du texte, il est essentiel d'avoir conscience de l'hétérogénéité – ainsi que du caractère nécessairement partiel et perfectible des réalisations – afin de pouvoir s'orienter au mieux dans sa quête d'outils utiles et aussi fiables que possibles ».

La mise en place d'analyses outillées doit donc reposer sur un équilibre entre les objectifs de l'analyse, les outils existants et le coût de la mise en œuvre de ces outils. Cet équilibre suppose donc un choix d'outils « éclairé » sur la connaissance de leur mode de fonctionnement et de leurs limites potentielles par rapport aux objectifs de l'analyse. En d'autres termes, il faut, dans la mesure du possible, non pas choisir un outil simplement parce qu'on en dispose, mais parce que l'on juge que son fonctionnement permet de faire ressortir des phénomènes linguistiques pertinents pour une analyse donnée et d'offrir par là-même des possibilités de description inédites.

Pour notre part, comme nous le décrivons dans la suite de ce chapitre, les outils sélectionnés sont des outils classiques souvent utilisés en terminologie, tels qu'un extracteur de termes, un concordancier et un extracteur de relations conceptuelles. Ce choix nous permet à la fois de placer le terme au centre de notre analyse pour observer ensuite son fonctionnement linguistique et les variations qu'il peut subir dans le système de la langue (à la manière des approches en terminologie textuelle, cf. Condamines & Rebeyrolle, 1997b ; Slodzian, 2000 : 73 au Chapitre 2), mais également d'utiliser des outils et des chaînes de traitement connus et répandus en terminologie. Ceci permet, d'une part, de situer notre méthode dans la lignée des méthodes d'analyses terminologiques classiques à partir de textes et, d'autre part, de considérer notre démarche comme peu coûteuse en temps d'apprentissage pour les linguistes/terminologues qui travaillent déjà en synchronie.

Maintenant que nous avons décrit les principaux éléments méthodologiques en jeu dans l'approche que nous proposons, nous illustrons plus précisément les choix qui s'imposent à

⁵⁵ En particulier des contraintes d'ergonomie et de gestion de corpus comparables. Assez peu d'outils en effet permettent de gérer aisément plusieurs sous-corpus à comparer simultanément.

l'analyste pour la constitution et le traitement des corpus, ainsi que le choix des outils pour aborder l'analyse de l'évolution. Pour ce faire, la section suivante détaille les deux corpus diachroniques sélectionnés pour notre étude et la manière dont ils ont été construits pour le domaine spatial

3.2 Choix méthodologiques pour notre étude

3.2.1 Corpus pour le domaine spatial

Comme nous l'avons présenté dans l'introduction de ce travail, notre recherche est née d'un besoin exprimé dans le domaine spatial. Ce besoin permet de nous guider dans l'ensemble des choix de constitution des données à observer, à commencer par celui des sources à sélectionner.

3.2.1.1 Choix des sources

Dans la mesure où nous inscrivons notre démarche dans la lignée de la terminologie textuelle, nous l'avons dit, les corpus de textes sont les sources privilégiées pour l'analyse proposée. Néanmoins, deux autres types de ressources peuvent être envisagés : des ressources terminographiques et le recours à des informateurs humains.

Le recours à des ressources terminographiques pour observer l'évolution est envisagé lorsque l'analyse se concentre sur l'évolution du lexique, comme cela est souvent le cas dans l'étude des langues de spécialité (Chapitre 1). Dans ce cas, on cherche à repérer les évolutions du lexique à partir de la comparaison de ressources produites à différentes périodes. Cette approche peut être illustrée par plusieurs types d'analyses. Le premier consiste à comparer plusieurs ressources lexicographiques de périodes différentes, par exemple pour observer et décrire l'évolution de ces ressources (voir par exemple des travaux tels que ceux de (Wesemaël & Wesemaël, 2005) sur l'évolution des dictionnaires Larousse entre 1955 et 2005). Le deuxième est la comparaison de ressources pour observer l'évolution d'une terminologie tel que le propose Møller (1994). Un troisième type d'approche met en œuvre la comparaison de ressources et de corpus de textes pour le repérage automatisé de néologismes (Drouin, *et*

al., 2006 ; Paquin, 2007). Comme décrit au Chapitre 1, il s'agit de comparer un corpus terminographique de référence avec des corpus de textes récents pour extraire les termes qui ne sont présents que dans les corpus de textes, termes qui constituent alors des néologismes potentiels (par exemple Cabré, *et al.*, 2003).

L'intérêt du recours à ces ressources pour observer l'évolution est lié au fait qu'elles permettent de se concentrer directement sur l'étude du lexique, lexique qui a été enregistré dans ces ressources comme représentatif de l'état de langue ou de connaissance au moment de leur conception. De plus, dans les cas où ces ressources associent à chaque entrée une définition, on peut envisager de comparer des évolutions de sens répertoriées. Néanmoins, une des difficultés majeures dans ce type de démarche est que ces ressources sont déjà une interprétation de la part de lexicographes sur un état de langue ou de connaissances donné, interprétation qui implique la prise en compte préalable de corpus de textes et, éventuellement, de décisions théoriques, linguistiques et éditoriales qui peuvent être difficiles à maîtriser pour construire un corpus comparable (à moins que l'on ne cherche précisément à étudier l'évolution de différentes générations d'une même ressource, comme le Larousse pour Wesemaël et Wesemaël (cités *supra*)). De plus, dans notre perspective, bien que le terme/concept soit central pour toute analyse terminologique, nous avons vu au chapitre précédent que le repérage de l'évolution doit plutôt chercher à bénéficier d'une analyse des langues de spécialité comme système et donc du terme dans les textes. Pour ces raisons, le recours systématique à des ressources terminographiques a été écarté, mais leur intérêt reste majeur. Pour cette raison, nous consultons ponctuellement certaines ressources⁵⁶ afin de mieux définir et traiter certains termes, mais, cependant, non à des fins de comparaison dans le temps.

Un autre type de ressource envisageable est le recours à des informateurs humains. La collaboration avec des experts fait écho au besoin d'interprétation lié à une approche sur corpus, tel que mentionné *supra* (§3.1.2.2, p.73), mais également à l'implication des experts dans les projets spatiaux, ainsi qu'à la volonté du Cnes d'impliquer ses experts dans les tâches de gestion des connaissances. Cependant, ce point de vue est très rarement – voire

⁵⁶ En particulier Termium (<http://www.btb.termiumplus.gc.ca/site/termium.php?lang=fra&cont=001>), le Grand Dictionnaire Terminologique (www.granddictionnaire.com) et le dictionnaire de spatologie (Cnes/Cilf, 1985 ; 1992 ; 1997 ; 2001).

jamais – proposé pour l'étude de l'évolution, la raison principale étant bien entendu le fait qu'il ne reste généralement plus de locuteurs ayant de compétence sur un état de langue ancien (Prévost, 2005 ; 2008). Cependant, dans le cas où l'analyse porte sur l'évolution des connaissances, il peut exister des experts (voire des historiens des sciences) disposant de compétences pertinentes et utiles pour décrire l'évolution, *a fortiori* lorsque la description se concentre sur des états de langue ou de connaissances contemporains. Quand on dispose d'experts on peut envisager de fonder l'analyse sur un échange avec eux. L'interaction qui en résulte est d'autant plus riche qu'elle repose sur les connaissances d'informateurs qui bénéficient d'une expérience et d'une pratique courante de la langue et/ou du domaine étudié(e), alors que le linguiste ne possède en général aucune compétence dans la langue de spécialité observée. Les informations fournies permettent de construire des analyses solides, voire incontestables, de l'évolution. De plus, travailler avec des experts permet de les impliquer dans des tâches qui peuvent aboutir à la mise en place d'applications dont ils sont eux-mêmes les utilisateurs, telles que la construction ou la mise à jour de ressources terminologiques (Aussenac-Gilles, *et al.*, 2002). En synchronie, dans le cas des langues de spécialité, il est d'ailleurs courant d'avoir recours à des experts du domaine dans différentes tâches terminologiques, telles que la validation des termes/concepts pour construire des ressources terminologiques (par exemple Bourigault & Lamé, 2002 ; Estopà Bajot, 2001). Nous verrons qu'en diachronie, il est beaucoup plus difficile d'interroger les experts dans la mesure où l'on questionne leur compétence *dans le temps*. Ce *rétro-diagnostic*, comme nous l'appellerons à la suite de Dury (2007), pose de nombreuses difficultés que nous détaillons dans le Chapitre 10. Mais pour ces différentes raisons, la collaboration avec les experts pour notre démarche est essentielle, bien que nous reviendrons dans la troisième partie de ce travail sur la question du rôle et de la place de l'expert dans une analyse linguistique diachronique de l'évolution.

3.2.1.2 Domaine d'activité et domaine de connaissance

Au vu de la difficulté à définir la notion de *domaine* en langue de spécialité (§2.1.1.3), une discussion de cette question dans le domaine spatial s'impose. Cette distinction conditionne en effet les choix faits dans notre recherche pour « situer » les corpus à sélectionner

(Condamines, 2003a ; Condamines, 2007) et assurer leur représentativité par rapport à notre problématique.

Nous l'avons vu, la répartition en domaines est souvent conventionnelle et définie *a priori*. Dans notre recherche, la répartition choisie est celle du Cnes dont l'activité repose sur un découpage propre au domaine spatial et aux besoins de l'organisme. Néanmoins, certaines caractéristiques propres à ce découpage doivent être mises en avant pour appréhender l'étude de l'évolution.

Le terme « domaine spatial » renvoie à la fois à l'ensemble des sciences et techniques spatiales, mais aussi au champ d'applications ayant trait à l'espace. Ce terme, derrière une apparence d'homogénéité, cache une hétérogénéité extrêmement forte du domaine spatial où se mêlent domaines scientifiques, domaines techniques, domaines dédiés au spatial, domaines appliqués au spatial, etc.

Parmi l'ensemble des conséquences que cette hétérogénéité peut avoir dans le cadre d'une étude linguistique, deux phénomènes principaux doivent être pris en compte pour traiter l'évolution :

1. bon nombre de domaines impliqués dans le spatial ont une existence en dehors du contexte spatial et s'inscrivent alors dans ce que l'on pourrait appeler une « double vitesse » ;
2. il est nécessaire de prendre en compte la distinction entre « domaine de connaissance » et « domaine d'activité » (Bessé (De), 2000).

Le premier point implique que certains domaines ne sont pas *dédiés* exclusivement au spatial, mais peuvent être *appliqués* au spatial. C'est l'exemple notamment de l'optique, domaine sur lequel porte en partie cette étude. L'optique est une discipline très ancienne puisque les premiers miroirs et l'utilisation du verre sont connus depuis l'Antiquité ; et les premières inventions d'instruments d'observation astronomique sont attribuées aux peuples khmers, chinois et mayas à partir de 2500 avant Jésus Christ. Cette discipline a depuis fait l'objet de recherches incessantes, jusqu'à aujourd'hui où elle connaît un essor fulgurant grâce à la photonique et à l'utilisation du laser et des fibres optiques (voir à ce sujet la thèse Lelubre (1992) sur l'histoire de la terminologie de l'optique en arabe) . Le spatial, quant à lui, est un

domaine récent développé à partir des années 50. Et c'est à partir de cette date que l'optique, depuis toujours attirée vers l'observation de l'espace, s'applique à l'environnement et aux besoins spécifiques du spatial : développement de télescopes, d'instruments d'observation de la terre et de l'espace, de technologies de communication, développement d'instruments répondant aux contraintes de précision, de taille, de poids des applications spatiales, etc. Ses apports et sa spécialisation pour le spatial sont tels que l'on parle aujourd'hui d'optique spatiale, conférant à cette dernière un statut de domaine à part entière. On serait donc aujourd'hui face à deux domaines : l'optique et l'optique spatiale. Or, dans les faits, le second est peut-être plus une spécialisation du premier qu'un domaine différent : l'optique spatiale est en grande partie l'application et l'adaptation de connaissances et techniques optiques aux besoins propres à l'environnement spatial. De fait, lorsque des évolutions ont lieu en optique spatiale plusieurs points de vue peuvent coexister. Parmi ces points de vue :

1. de nouveaux composants utilisés en optique spatiale sont connus depuis longtemps en optique « générale », mais on ne sait les appliquer au spatial que depuis peu,
2. des techniques sont propres au spatial et ne se répercutent pas ou peu en optique,
3. des évolutions en optique ne sont pas utilisées dans le spatial.

Ce « double niveau » doit donc être pris en compte dans les interprétations de l'évolution du domaine observé. En effet, selon que l'on pose l'optique spatiale comme domaine à part entière ou comme spécialisation de l'optique, l'évolution observée peut être interprétée de manière sensiblement différente. Par exemple, en fonction du point de vue adopté, un concept optique nouvellement utilisé dans le spatial peut être interprété comme ancien s'il est connu depuis longtemps en optique, ou encore comme ancien mais ayant évolué parce qu'appliqué depuis peu au spatial, ou même nouveau si l'on s'accorde sur le fait que l'optique spatiale est un domaine à part entière. Ceci montre d'emblée la variabilité des interprétations auxquelles on peut faire face dans le cas de l'évolution des connaissances du domaine spatial. Mais nous y reviendrons longuement à l'aide d'exemples dans la suite de ce travail.

Le second point à souligner est la dimension technique centrale et l'importance du développement d'instruments, d'applications et de la mise en place de projets appliqués dans le domaine. Les développements techniques impliquent de prendre en compte non seulement les domaines de connaissances (au sens de De Bessé (2000 : 184)), mais également les domaines d'activité ou d'application. Rappelons que De Bessé (*ibidem*) définit un domaine d'activité comme

« permet[tant] d'identifier un champ d'action, un ensemble d'actes coordonnés, une activité réglée, une pratique. Il correspond à une activité humaine, sociale, économique, industrielle. Il est constitué d'un ensemble de procédés bien définis destinés à produire certains résultats ».

Cette définition, bien qu'elle reste assez vague, peut s'appliquer au cas des projets spatiaux : les projets spatiaux, qui visent la mise en place de missions, le développement d'instruments, la mise en œuvre de systèmes spatiaux, constituent le fondement de l'activité du Cnes. Ils demandent la mise en œuvre de différents domaines et métiers, de moyens humains et financiers spécifiques pour parvenir à l'objectif d'un projet précis. Et comme dans le cas de n'importe quelle entreprise où les activités terminographiques peuvent être « orientées et délimitées en fonction [...] d'une activité destinée à produire un résultat » (*ibidem*), les activités terminographiques peuvent, voire doivent, être orientées en fonction du domaine d'activité que sont les projets spatiaux.

Pour comprendre et saisir l'évolution dans le domaine spatial, il faut donc prendre en compte ces différents éléments et observer à la fois l'évolution des domaines de connaissances et l'évolution des domaines d'activité. Ces aspects conditionnent en effet en grande part la représentativité des corpus et, de fait, les interprétations que l'on peut proposer de l'évolution, celle-ci étant relative au domaine sélectionné pour l'analyse.

3.2.1.3 Contraintes de constitution des corpus diachroniques

Les corpus diachroniques doivent répondre à diverses contraintes pour assurer leur représentativité et leur pertinence, tel que le soulignent Condamines *et al.* (2004). À la suite de leurs travaux menés au Cnes, nous avons défini pour notre recherche six contraintes

principales afin de garantir le fait que les variations observées soient bien attribuables à une évolution dans le temps et non pas à un autre type de variation :

1. **Contrainte de diachronicité** : Les corpus construits doivent correspondre à des périodes clés dans le domaine spatial pour lequel est menée cette étude et à des intervalles courts (comparables à ceux rencontrés dans le cadre de projets spatiaux) afin d’observer l’évolution en diachronie courte. Ce point renvoie en grande part aux éléments présentés au §3.1.1.2 *supra*.
2. **Contrainte d’homogénéité du genre et du degré de spécialisation** : Les textes choisis doivent appartenir à un même genre et à un même degré de spécialisation pour neutraliser l’influence que ceux-ci pourraient avoir sur les variations repérées dans le corpus⁵⁷.
3. **Contrainte de variété des rédacteurs** : Dans la mesure du possible, les textes qui constituent les corpus doivent être écrits par plusieurs auteurs différents afin de circonscrire au maximum les variations particulières à un seul individu (idiosyncrasie).
4. **Contrainte de recours à des locuteurs natifs** : Les productions de rédacteurs natifs de la langue étudiée sont privilégiées afin d’éviter que les phénomènes observables soient dus à un manque de compétence langagière du rédacteur.
5. **Contrainte de disponibilité des sources** : Les textes choisis pour composer les corpus doivent provenir du Cnes, être accessibles et non protégés par le secret industriel.
6. **Contrainte de disponibilité d’un expert** : Dans la mesure où la collaboration avec des experts du domaine est primordiale dans notre recherche, pour chacun des corpus et des domaines qu’ils couvrent, des experts doivent être disponibles et accepter de nous aiguiller dans la description de l’évolution.

Ces contraintes sont strictes et font de la constitution des corpus une tâche ardue et de longue haleine. Il est de fait très difficile de satisfaire toutes ces contraintes. Néanmoins, avec l’aide

⁵⁷ Notons cependant que le genre est susceptible lui aussi d’évoluer dans le temps, mais nous ferons l’hypothèse qu’il restera stable en diachronie courte.

du Centre de Documentation du Cnes, nous avons sélectionné deux corpus comparables qui les satisfont pleinement :

- un corpus de trois éditions du cours de Techniques et Technologies des Véhicules Spatiaux,
- un corpus de rapports de spécification des première et troisième générations des balises DORIS (Doppler Orbitography and Radiopositionning Integrated by Satellite ou Détermination d'Orbite et de Radiopositionnement Intégrés par Satellite).

Nous les présentons dans la section suivante et détaillons la manière dont ils répondent à ces six contraintes. En guise d'illustration, des extraits de chacun de ces deux corpus sont donnés en Annexe A (p.377*sqq.*).

3.2.2 Description des corpus sélectionnés

3.2.2.1 Corpus TTVS

Le premier corpus constitué est un corpus de cours de Techniques et Technologies des Véhicules Spatiaux (désormais TTVS) rédigés par des experts du Cnes. Le but de ce cours est

« de présenter l'ensemble des techniques et des technologies qui concourent à la constitution et au fonctionnement des véhicules spatiaux.

Il présente les caractéristiques et les contraintes de conception spécifiques à la technologie spatiale et décrit les techniques et les technologies des sous-systèmes et des équipements embarqués. Il s'adresse plus particulièrement aux ingénieurs débutants ou aux ingénieurs confirmés qui souhaitent élargir l'éventail de leurs connaissances pour acquérir une vision technique plus globale.

Les trois volumes fournis constitueront une référence pour tous ceux qui exercent une activité technique dans le domaine spatial. »
(<http://cnes.cborg.net/ttvs2009/>, consulté le 11 novembre 2008)

Le TTVS est rédigé depuis 1994 par plus de 80 experts du Cnes appartenant à l'ensemble des spécialités métiers qui y sont présentes. Il s'agit du support écrit en 3 volumes d'un cours annuel de 10 jours organisé par le Cnes, disponible en librairie et publié par le Cnes et les

éditions Cépaduès. Ce cours est écrit par des experts à l'attention d'experts ou de semi-experts (Bowker & Pearson, 2002).

Si les sessions de cours sont annuelles, les trois volumes sont quant à eux publiés tous les 4 ans. Il s'agit là d'un aspect intéressant de ce corpus en lien avec sa périodisation dans la mesure où les intervalles de temps qui existent entre chaque sous-corpus⁵⁸ n'ont pas été choisis arbitrairement, mais émanent directement des besoins du domaine et de son rythme d'évolution :

« Voici l'édition 2002 du cours de Techniques et Technologies des Véhicules Spatiaux. Le cycle de 4 ans se trouve ainsi institutionnalisé (94 – 98...) et le lien avec le cycle de vie de nos développements technologiques n'est pas fortuit. [...] En quatre ans, des changements profonds de stratégie de développement des segments spatiaux sont intervenus. » (préface de D. Assémat (Cnes, 2002))

Le fait que le choix de l'intervalle et du découpage en périodes soit basé sur des critères externes identifiés par des experts du domaine revêt un intérêt particulier dans la mesure où, du point de vue de l'évolution des connaissances, ce type de découpage reflète une évolution ressentie par les acteurs du domaine eux-mêmes et dont la pertinence est le fruit d'un consensus entre eux.

L'autre aspect intéressant de ce corpus repose sur le fait que la diachronicité des textes est basée sur des rééditions d'un même support. Dans la mesure où chacune des rééditions est basée sur l'édition précédente, le corpus est construit selon une certaine « intertextualité », définie par Kristeva (1969) comme une interaction à l'intérieur d'un texte qui permet de saisir « les différentes séquences (ou codes) d'une structure textuelle précise comme autant de *transforms* de séquences (de codes) prises à d'autres textes ». Cette notion littéraire (développée entre autres par Barthes et Genette) n'est reprise ici que très partiellement et approximativement pour souligner le fait que chacune des éditions du cours se positionne par rapport aux précédentes. Certaines séquences ou commentaires présents dans les éditions les plus récentes trouvent alors un écho dans les éditions les plus anciennes. Cette propriété

⁵⁸ À partir de maintenant, nous nommerons « corpus » l'ensemble des textes qui composent le corpus comparable et « sous-corpus » les textes qui correspondent à une période/date donnée dans ce corpus. Ainsi, nous traiterons du corpus TTVS, lui-même composé de trois sous-corpus (TTVS1994, TTVS1998 et TTVS2002).

est intéressante pour aider à construire une interprétation de l'évolution puisque, comme nous le verrons, elle est susceptible de rentrer en jeu dans la comparaison des différentes éditions.

Enfin, bien que ce corpus réponde à l'ensemble des contraintes définies pour leur constitution, la diversité des domaines de connaissances abordés dans ce cours impose de restreindre notre recherche à l'un des domaines présentés dans le TTVS. En effet, les éditions du TTVS présentent en quatorze chapitres l'ensemble des techniques et technologies spatiales mises à contribution au Cnes : aperçu des principales missions spatiales, composantes du système spatial et principales étapes du développement d'un véhicule spatial, lois fondamentales de la mécanique spatiale, contraintes de l'environnement spatial, effets des rayonnements, assurance produit et développement des systèmes orbitaux, charges utiles (télécommunications, observation, sciences, localisation-navigation), missions et charges utiles de radiolocalisation et radionavigation, physique de la mesure, caractéristiques de l'instrumentation optique et infrarouge et de l'instrumentation radiofréquence, techniques radioélectriques, techniques optique et technologies des détecteurs et des électroniques de détection, etc. (Cnes, 2002). La diversité des domaines et questions abordés dans ce cours introduit une trop grande complexité pour observer l'évolution des connaissances et il s'avère nécessaire de concentrer nos observations sur un seul domaine.

Pour circonscrire le domaine observé dans cette recherche, nous avons privilégié les domaines pour lesquels des experts étaient disponibles. Dans la mesure où, au moment de cette recherche, le Cnes travaillait avec des experts sur la mise en place d'une ontologie de l'optique et optoélectronique spatiales, notre choix s'est porté sur ce domaine. À l'aide d'un expert en optique qui a lui-même participé à la rédaction du TTVS, certaines sections des chapitres 5 et 7 et le chapitre 9 dans son intégralité ont été sélectionnées pour chacune des trois éditions afin de construire un corpus d'optique spatiale. La taille du corpus et de chacun des sous-corpus est présentée dans le Tableau 3.1. Au total, quatre experts d'optique ont accepté de participer à notre recherche, tous ayant contribué à au moins l'une des éditions du TTVS.

TTVS1994	TTVS1998	TTVS2002	TTVS (Total)
46 448 occurrences	78 656 occurrences	109 505 occurrences	234 609 occurrences

Tableau 3.1 – Nombre d'occurrences – Corpus et sous-corpus TTVS

À la lecture de ce tableau, une remarque doit être faite. Les trois sous-corpus apparaissent clairement déséquilibrés : la taille des sous-corpus augmente au fur et à mesure des éditions et le sous-corpus TTVS1994 est presque 3 fois plus petit que le TTVS2002. Il s'agit là d'une des difficultés majeures pour le traitement de ce corpus. Cependant, bien que cela entraîne un déséquilibre certain, ce phénomène est révélateur d'évolutions dans le domaine : si celui-ci progresse, que de nouvelles techniques ou technologies voient le jour, des modifications viennent compléter le cours d'édition en édition. Néanmoins, une difficulté est que les auteurs peuvent également rectifier des erreurs, compléter des oublis ou autres lors de rééditions. Dans ce cas, il ne s'agit pas d'une évolution du domaine pertinente pour notre recherche, mais plus d'une évolution du cours en tant que manuscrit. Pour essayer de pallier ce phénomène, une possibilité aurait été de redécouper ce corpus pour l'équilibrer. Pour ce faire, plusieurs possibilités existent, telles que tronquer arbitrairement les sous-corpus pour que chacun contienne au plus 45 000 occurrences, faire une sélection aléatoire des textes ou des segments à conserver sur la base de calculs statistiques, etc. Nous avons envisagé chacun de ces cas, mais ce type de manipulations entraîne une perte importante d'informations sur l'évolution. On perd en particulier des informations sur la structure du texte en tant que discours construit et ainsi certaines sections « stratégiques » dans les textes qui contiennent beaucoup de connaissances évolutives (si l'on tronque la fin des textes : conclusions, sections de perspectives, comparaison des résultats, etc. et si l'on tronque le début : introductions, états de l'art, etc.). Au cours de nos recherches, nous avons cependant travaillé sur une version équilibrée du corpus DORIS obtenue par comparaison des sections stables d'une édition à l'autre (Picton, 2007). Le tri manuel que nous avons effectué a été guidé par le logiciel KDiff⁵⁹. Au vu des limitations que l'équilibrage entraîne, nous revenons pour cette recherche à une exploration du corpus dans son ensemble.

⁵⁹ Développé par J. Eibl, <http://kdiff3.sourceforge.net/doc/index.html> (consulté le 10 novembre 2008) et conçu à l'origine pour aligner des corpus parallèles.

Un second corpus est choisi pour cette étude, constitué de textes du projet spatial DORIS. En effet, si le corpus TTVS se veut un corpus de connaissances scientifiques et techniques sur un court intervalle, nos observations doivent être complétées sur un « corpus projet ».

3.2.2.2 Corpus DORIS⁶⁰

Le projet DORIS est un projet d'élaboration d'un système d'orbitographie conçu et développé par le Cnes en collaboration avec le GRGS (Groupe de Recherches de Géodésie Spatiale) et l'IGN (Institut Géographique National). Ce système est

« utilisé pour déterminer au centimètre près l'orbite de satellites équipés de récepteurs Doris grâce à un réseau de stations terrestres, utilisées comme points de référence au sol. À l'inverse, le système permet le rattachement précis de points donnés au Système de référence terrestre international (ITRF). » (<http://www.cnes.fr/web/CNES-fr/869-doris.php>, consulté le 11 novembre 2008)

Le projet DORIS est un projet de longue durée qui a démarré à la fin des années 80 et est toujours en cours actuellement. Les premières stations DORIS au sol ont été déployées entre 1987 et 1990. Le premier récepteur a été lancé à bord du satellite Spot en 1990. Au milieu des années 90, une deuxième génération du système a vu le jour, puis fin 90, la troisième génération a été mise en place. Aujourd'hui, la quatrième génération est en cours d'élaboration et plusieurs lancements sont prévus entre 2009 et 2011, soit plus de 20 ans après le début du projet. Depuis le début du projet, l'instrument DORIS a été embarqué sur de nombreuses missions Cnes, telles que Topex-Poséidon, Spot 2 à 5, ENVISAT, Jason1, Jason2, etc.

Ce projet est intéressant dans la mesure où son évolution repose sur une succession de générations d'instruments qui sont le résultat entre autres d'évolutions techniques, telles que la miniaturisation, des nouvelles possibilités de mesure, etc. Il s'agit donc d'un contexte idéal

⁶⁰ Le corpus que nous utilisons est une version redécoupée du corpus DORIS utilisé lors d'un projet intitulé « Méthodes et outils de *data mining* pour les projets spatiaux » mis en place au Cnes (2002-2004). Dans ce projet, le laboratoire CLLE-ERSS a posé les premières bases de ce que pourrait être une méthode de repérage de l'évolution à partir de textes (Condamines *et al.*, 2004). Nous présentons ces travaux plus loin dans notre étude (notamment §7.3.4).

pour apprendre à repérer linguistiquement les évolutions de connaissances en jeu dans un projet.

Dans un projet tel que DORIS, la documentation technique est nombreuse et variée : dossiers de définition, manuels utilisateur (ou maintenance), plans (de gestion, d'essai, planning), rapports (d'études, d'essais, d'avancement), budgets, contrats (appels d'offres, rapports de choix, de présentation, contrat, clauses techniques, doc financiers), cahiers des charges fonctionnel, documents de contrôle des interfaces, compte-rendus de réunion, dossiers industriel, demandes de modification, demandes d'utilisation, demandes de dérogation, fiches d'anomalie, listes (composants, matériaux, documentation, actions, tâches), livrets suiveur, notes techniques, procédures, registres de contrôle individuel, copies papier des transparents, demandes de déviation , etc.

Parmi ces documents, tous n'interviennent pas aux mêmes étapes du projet et surtout, tous appartiennent à des genres différents. Comme dans le cas du TTVS, il faut s'assurer de construire un corpus homogène dont on puisse dire que les variations présentes sont dues à une évolution du projet. Pour ce faire, le corpus construit pour cette recherche reprend les réflexions proposées par Condamines, Rebeyrolle et Soubeille (2004) dont les analyses étaient basées sur le projet DORIS. Les textes sélectionnés à l'époque étaient les rapports de spécification (c'est-à-dire les cahiers des charges) des trois premières générations de balises construites depuis le début du projet, documents que nous reprenons pour cette recherche. Ces derniers sont écrits par différents ingénieurs mobilisés sur le projet DORIS et ce, pour chacune des trois premières générations. Néanmoins, à la différence des travaux de Condamines *et al.* qui se basaient sur l'ensemble des trois générations, les textes de la deuxième génération de balise sont exclus de notre analyse. Un nouveau corpus est reconstitué à partir des documents disponibles pour les première et troisième générations uniquement. En effet, la seconde génération est un peu particulière : comme la première et la troisième, l'application visée par le projet est la localisation des satellites et de points précis où sont placées les stations au sol. Mais les balises de deuxième génération visent en plus une application « de terrain » pour localiser de nouveaux points tels que le déplacement d'un glacier, un volcan, etc. où il n'y a pas d'infrastructure DORIS sur place. Pour ce faire, on a recours à des balises d'appui en laboratoire, et la balise DORIS est développée différemment

pour répondre à ces nouvelles contraintes de terrain : son boîtier est plus petit et consomme moins d'électricité, elle est équipée de batteries et panneaux solaires, etc. Cette application « de terrain » a été retirée des balises de troisième génération pour revenir à des objectifs plus proches de la première. Pour cette raison, ne seront comparés ici que des rapports de spécification des première et troisième générations.

Le corpus DORIS final est donc composé de 5 rapports de spécifications de la première génération et 6 rapports de spécification de la troisième génération, soit un total de 36 401 occurrences (Tableau 3.2).

Ce corpus est un très petit corpus, équilibré et idéal pour une exploration fine et qualitative des phénomènes d'évolution.

DORISGEN1	DORISGEN3	DORIS (Total)
17 544 occurrences	18 857 occurrences	36 401 occurrences

Tableau 3.2 – Nombre d'occurrences – Corpus et sous-corpus DORIS

Le Tableau 3.3 reprend les principales caractéristiques des deux corpus retenus par rapport aux contraintes identifiées *supra* en 3.2.1.3 (p.82).

		TTVS	DORIS
Diachronicité	Intervalle	8 ans	11 ans
	Période	1994-1998-2002	1989-2000
	Découpage	3 sous-corpus découpés tous les 4 ans	2 sous corpus (1989 et 2000)
Genre		Cours édités et proposés en vente libre Didactique	Rapports de spécification (cahiers des charges) Technique
Textes		1 chapitre entier et plusieurs sections de deux autres chapitres pour chacun des 3 sous-corpus	5 rapports de spécification de la première génération (1989) 6 rapports de spécification pour la troisième génération (2000)
Domaines		Optique et optoélectronique → domaine de connaissance	Techniques et constitution de l'instrument DORIS → domaine d'activité (corpus projet)
Variété des rédacteurs		Plus de 80 rédacteurs ingénieurs au Cnes En moyenne 10 rédacteurs par chapitre du cours	1 rédacteur pour chaque rapport de spécification Rédacteurs différents pour la quasi-totalité des rapports Une dizaine de rédacteurs en tout
Degré de spécialité		Expert à semi-expert	Expert à expert
Locuteurs natifs		Majorité des auteurs	Tous
Disponibilité des sources		Sous droits d'éditeur (Cnes et Cépaduès Éditions)	Textes Cnes du projet DORIS
Experts disponibles		4 experts	1 expert
Nombre d'occurrences		234 609	36 401

Tableau 3.3 – Description des corpus en fonction des contraintes de constitution

3.2.3 Choix des outils et traitement des corpus

Nous présentons dans cette section les outils choisis et la manière dont les données ont été préparées pour notre analyse, en regard avec notre position dans une approche linguistique outillée.

3.2.3.1 Choix des outils

Soulignons succinctement le fait que les corpus obtenus ont dû être préparés et « nettoyés » pour pouvoir être exploités par les outils :

- Le TTVS2002 était disponible au format numérisé (*.doc). Il a simplement été converti au format texte. Les deux premiers sous-corpus (TTVS1994 et TTVS2002) n'étaient disponibles qu'au format papier. Ils ont donc été scannés et ocrisés pour obtenir des formats texte exploitables⁶¹. Chacun des sous-corpus TTVS a été relu et corrigé manuellement pour éliminer les erreurs de numérisation et de conversion des caractères spéciaux au format texte, ainsi que pour traiter les figures, tableaux et formules présents dans le corps du texte (nous avons fait le choix de supprimer les tableaux de chiffres, formules et figures, notés respectivement TABL, FIGR et FRML pour garder une trace de ces éléments lors de l'exploration des textes).
- Dans la mesure où le corpus DORIS avait été constitué dans le cadre d'un projet antérieur, toutes les conversions et corrections avaient été effectuées. Les textes étaient donc prêts à être compilés.

Comme nous l'avons dit plus haut, la démarche outillée que nous préconisons s'inscrit dans la lignée de celle définie dans le cadre de la terminologie textuelle. De fait, les termes sont considérés comme les pivots de notre analyse et constituent le point d'entrée sur lesquels s'appuie l'analyse. La description est menée à l'aide d'outils classiques en terminologie textuelle : un extracteur de termes, un extracteur de relation et un concordancier. Plus précisément, les outils choisis sont l'analyseur Syntex, l'outil TerminoWeb et le concordancier AntConc, choix que nous justifions dans les paragraphes suivants.

A. Syntex

Syntex est un analyseur syntaxique de corpus développé par Bourigault *et al.* (2007 ; 2000 ; 2005), qui permet d'extraire d'un corpus une liste de candidats termes et de syntagmes, structurée par des relations de dépendance syntaxique. L'intérêt de l'outil pour cette recherche est triple.

Le premier atout de Syntex pour cette recherche repose sur la souplesse de l'outil. En effet, cet analyseur est conçu pour s'adapter à différents besoins (Bourigault, 2007 ; Bourigault, *et al.*, 2005), qu'il s'agisse d'applications en terminologie (en particulier pour la construction de

⁶¹ Cette lourde étape n'aurait pas été possible sans l'aide du centre de documentation du Cnes et en particulier de Daniel Galaretta et de Christine Gabens, que nous remercions.

ressources termino-ontologiques à partir de textes et la structuration de termes (Aussenac-Gilles & Bourigault, 2003 ; Bourigault, *et al.*, 2004)) ou de recherches descriptives en corpus (par exemple Fabre & Bourigault, 2008). L'outil Syntex s'impose donc pour répondre à nos objectifs, à la fois descriptifs et appliqués.

Le second intérêt de l'outil est lié à son mode de fonctionnement qui repose sur une procédure « d'apprentissage endogène », c'est-à-dire sans aucune autre ressource que les seules données du corpus. Cette propriété est particulièrement intéressante en diachronie dans la mesure où elle permet d'éviter d'avoir à contrôler et gérer la périodicité de ressources externes. En d'autres termes, l'apprentissage endogène permet de concentrer l'analyse diachronique sur le corpus construit et de ne pas intégrer d'autres données qui pourraient venir brouiller la cohérence diachronique du corpus. De la même manière, puisque le fonctionnement endogène de Syntex implique que l'outil n'intègre pas de connaissances de la terminologie du domaine *a priori*, le traitement peut être réappliqué à n'importe quels domaines et périodes pour lesquels on dispose de corpus.

Enfin, le troisième intérêt de cet outil repose sur le fait que, contrairement à la plupart des outils terminologiques du même type, Syntex fait porter son analyse sur toutes les unités lexicales et non pas uniquement sur les noms. Ce point de vue s'éloigne de la vision terminologique classique du terme comme appartenant à la catégorie nominale exclusivement et permet d'intégrer des perspectives nouvelles. L'Homme (2002 ; 1998) souligne en particulier l'importance des verbes pour la description terminologique et montre que ceux-ci peuvent être considérés soit en tant qu'unités terminologiques à part entière soit comme des supports centraux pour analyser le sens des termes et aider à construire des classes de termes.

D'un point de vue ergonomique, grâce à l'interface TermOnto (voir Baneyx, *et al.*, 2005 par exemple pour une utilisation de TermOnto), les analyses Syntex sont facilement consultables. Cette interface est développée sous Access par Didier Bourigault mais nous l'avons légèrement adaptée pour prendre en compte des corpus comparables⁶². Afin de mieux

⁶² Comme souligné *supra*, de manière générale, les outils disponibles manipulent difficilement les corpus comparables et leurs fonctionnalités sont souvent peu ergonomiques, obligeant l'analyste à ouvrir autant de

illustrer les fonctionnalités de Syntex et la visualisation, les figures suivantes présentent des copies d'écran des sorties d'analyses Syntex dans TermOnto.

The screenshot displays the Microsoft Access - [Liste] window, which serves as the interface for TermOnto. It shows a list of candidate terms with their syntactic analysis results. The interface is organized into a grid-like structure with multiple sections for each candidate term.

Candidate 1: laser

- Nom:** laser
- Validité:**
 - ☒ Terme du domaine
 - ☐ Je ne sais pas
 - ☐ Pas terme
- Années:**
 - ☐ 1994
 - ☐ 1998
 - ☐ 2002
- Contextes:**

94	98	02
772	140	290
...
- Commentaires:**
- productivité:** 135
- nb voisins:** 120
- var:** 11
- nbdoc:** 137

Candidate 2: détecteur

- Nom:** détecteur
- Validité:**
 - ☒ Terme du domaine
 - ☐ Je ne sais pas
 - ☐ Pas terme
- Années:**
 - ☐ 1994
 - ☐ 1998
 - ☐ 2002
- Contextes:**

94	98	02
752	172	237
...
- Commentaires:**
- productivité:** 116
- nb voisins:** 270
- var:** 278

Candidate 3: optique

- Nom:** optique
- Validité:**
 - ☒ Terme du domaine
 - ☐ Je ne sais pas
 - ☐ Pas terme
- Années:**
 - ☐ 1994
 - ☐ 1998
 - ☐ 2002
- Contextes:**

94	98	02
727	113	217
...
- Commentaires:**
- productivité:** 7
- nb voisins:** 269
- var:** 294

Candidate 4: sional

- Nom:** sional
- Validité:**
 - ☐ Terme du domaine
 - ☐ Je ne sais pas
 - ☐ Pas terme
- Années:**
 - ☐ 1994
 - ☐ 1998
 - ☐ 2002
- Contextes:**
- Commentaires:**
- productivité:**
- nb voisins:**
- var:**
- nbdoc:**

The bottom of the window shows a status bar with the text "Enr : 1 sur 2558 (Filtré)" and "Mode Formulaire".

Figure 3.1 – Extrait de la liste Syntex de candidats-termes – Interface TermOnto

Les résultats Syntex sont présentés sous forme de listes de candidats termes, auxquels sont associées des informations sur la fréquence dans le corpus et, dans chacun des sous-corpus, la classe syntaxique du candidat (Figure 3.1). De plus, un lien hypertextuel permet d'accéder à une fiche où sont synthétisées certaines caractéristiques du comportement linguistique du candidat dans les corpus examinés (dont les dépendances syntaxiques) sur lesquels fonder l'analyse (Figure 3.2).

fenêtres d'interface que de sous-corpus traités. Les transformations apportées à l'interface TermOnto permettent de limiter certaines de ces difficultés.

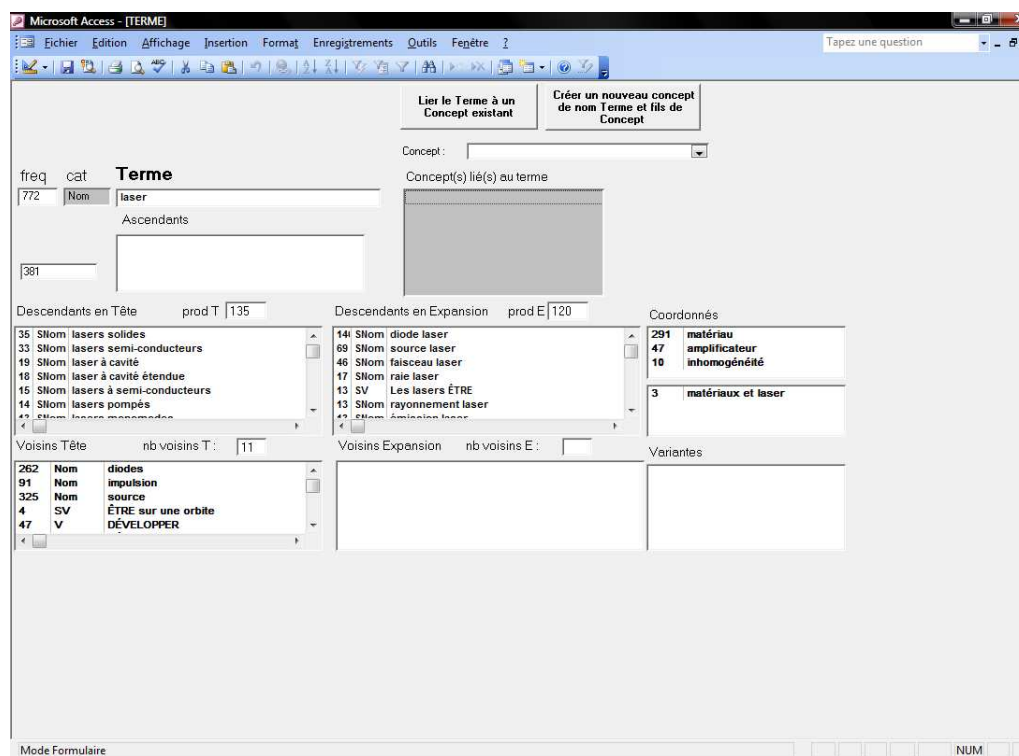


Figure 3.2 – Extrait de l'analyse syntaxique des dépendances (TermOnto)

Néanmoins, bien que les informations fournies par Syntex soient très riches, l'analyste de corpus ne peut se passer d'un outil plus souple de type concordancier pour naviguer dans les textes. Pour cette raison, nous complétons la description par l'utilisation du concordancier AntConc.

B. AntConc

AntConc est un logiciel libre⁶³ développé par Anthony (2005). Notre choix s'est porté sur ce concordancier pour plusieurs raisons. La première est qu'il s'agit d'un logiciel gratuit, facile d'installation et d'utilisation. Il permet de visualiser les termes en contextes, mais également de construire des expressions régulières très fines pour affiner les recherches, de calculer les cooccurrences, etc. Une des particularités de cet outil repose également dans ses possibilités de traiter aisément différents sous corpus : il permet de compter et visualiser les occurrences d'une unité recherchée dans chaque sous-corpus comme dans le corpus entier, propriété intéressante pour une comparaison en diachronie.

⁶³ Le concordancier est disponible à l'adresse : <http://www.antlab.sci.waseda.ac.jp/software.html> (consulté le 3 avril 2009)

C. TerminoWeb

Enfin, le dernier outil utilisé est le logiciel TerminoWeb, développé par Barrière (2006). Cet outil est développé pour assister la construction des ressources termino-ontologiques. Les différentes fonctionnalités proposées comprennent la construction semi-automatique de corpus, l'extraction de termes dans les corpus, le repérage de variantes terminologiques, l'aide à la structuration des ressources et l'exploration de relations sémantiques (taxonomiques et non-taxonomiques). C'est précisément pour cette fonctionnalité que cet outil a été choisi dans la mesure où l'une de nos hypothèses est qu'il est possible de définir des marqueurs d'évolution dans les textes pour repérer des contextes riches en connaissances évolutives (Chapitre 5). Généralement, les outils d'extraction d'informations à partir de marqueurs et patrons linguistiques contiennent des listes prédéfinies de marqueurs et sont difficilement modifiables. L'intérêt de TerminoWeb pour notre recherche est qu'il nous permet de définir notre propre liste de marqueurs, de la modifier et de la projeter dans les textes afin de comptabiliser et de visualiser les contextes extraits *via* son interface (Figure 3.3). De plus, cet outil gère très bien les corpus comparables.

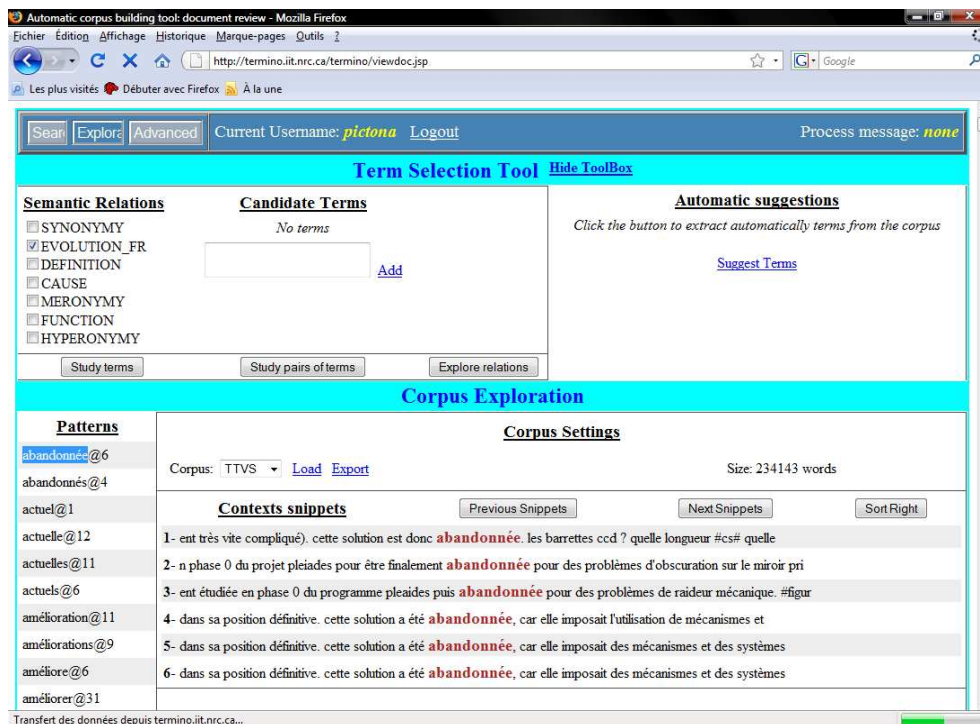


Figure 3.3 – Interface de visualisation (TerminoWeb)

Notons enfin, qu'en plus de ces trois outils, les analyses sont complétées par des programmes Perl ponctuels développés par nos soins ou adaptés de scripts disponibles sur Internet (notamment pour le calcul du χ^2 , Chapitre 4).

Après avoir décrit les étapes de constitution des corpus, ainsi que les outils pour les exploiter, il convient de présenter le traitement et l'extraction des données en corpus effectués pour l'analyse, et notamment l'extraction des termes à partir desquels nous commençons l'analyse.

3.2.3.2 Extraction et préparation des données

La première étape d'analyse dans notre démarche consiste en l'extraction des termes avec l'outil Syntex.

À partir de corpus étiquetés et lemmatisés, Syntex procède à une analyse syntaxique en dépendances et fournit en sortie une liste de syntagmes (nominaux, verbaux, adjectivaux et adverbiaux) et de mots catégorisés (noms, verbes, adjectifs, adverbes et noms propres) présentés comme « candidats-termes ». L'analyse Syntex seule produit cependant beaucoup de bruit qu'il faut filtrer avant de parvenir à une liste de termes acceptable. Ainsi, dans le corpus TTVS, l'analyse Syntex propose en sortie une liste de plus de 45 000 candidats-termes (pour un corpus de 230 000 occurrences) et de plus de 8 200 pour le corpus DORIS (qui compte 36 000 occurrences). Une sélection s'avère donc nécessaire. L'observation de la liste obtenue révèle en effet de nombreuses erreurs telles que des candidats trop longs, des erreurs d'analyses syntaxiques, des mots trop généraux, etc. Le bruit provoqué par l'extraction automatique des termes est très souvent mis en avant par les terminologues dont L'Homme (2001) ou Carreño *et al.* (à paraître). Ces derniers comparent d'ailleurs les résultats de plusieurs extracteurs et soulignent que Syntex est l'un des outils qui propose les listes les plus longues et bruitées.

Pour pallier ce phénomène et obtenir une liste satisfaisante, nous avons établi une série de trois filtres à partir de critères de sélection des termes proposés notamment par Delavigne (2001b) et L'Homme (2004). Ces critères reposent essentiellement sur : la fréquence du candidat, sa catégorie grammaticale, sa forme syntaxique et son appartenance au domaine.

Dans un premier temps, la sélection des candidats-termes (désormais CT) est filtrée automatiquement à partir d'un programme Perl que nous avons développé. Ce filtre fait intervenir le critère de forme syntaxique et permet d'éliminer :

- les CT qui contiennent les auxiliaires « être » et « avoir » (*exigence être, température être suffisante, avoir du vignettage*) ;
- les CT qui contiennent des chiffres arabes ou romains, des unités de mesures, des symboles mathématiques, des signes de ponctuation – à l'exception du point que l'on peut trouver dans certains sigles (par exemple : *résolution de 10 m, paragraphe VII.2.5.3.3, CAN de 8*) ;
- les CT trop longs, dont la taille a été fixée à plus de quatre unités lexicales après observation des listes⁶⁴ (par exemple : *le satellite SPOT assure une résolution au sol de 10 m avec une fauchée de 60 km, Description simplifiée de l'instrument adapté à une prise de vue globale*) ;
- les CT qui contiennent les marques de formule (FRML), tableau (TABL), et figure (FIGR) notées lors de l'étape de nettoyage des corpus (par exemple : *fréquence spatiale FRML, conductivité thermique TABL*).

C'est également à ce stade que sont choisies les catégories grammaticales conservées pour l'analyse. Et au vu de ce qui a été souligné *supra* (§3.2.3.1, p.91), l'ensemble des catégories étiquetées par Syntex sont prises en compte : dans la mesure où l'intérêt des catégories verbales et adjectivales a été montré pour la terminologie, il est important de les observer pour étudier le changement.

Dans un second temps, un filtre de fréquence est appliqué : seuls les CT qui apparaissent au moins quatre fois dans l'ensemble du corpus TTVS et trois fois dans le corpus DORIS sont conservés, ce qui permet d'éliminer les CT de trop basse fréquence (par exemple *inertie en flexion, outillage tournant, contraintes photométriques*, etc.). Nous choisissons un seuil de fréquence volontairement bas pour deux raisons principales :

- les corpus analysés comptent peu d'occurrences et beaucoup de termes n'affichent de fait qu'une faible fréquence,

⁶⁴ Cette limite reste néanmoins arbitraire.

- la liste souhaitée n'est pas constituée à des fins terminographiques, mais pour servir de base à la mise en œuvre des indices. Et dans la mesure où les termes évolutifs ne sont pas connus *a priori*, l'objectif de cette sélection est avant tout d'éliminer les termes trop peu fréquents pour que l'on puisse y appliquer les indices.

Enfin, à partir de cette sélection, le troisième filtre appliqué est un filtre manuel pour éliminer les erreurs non repérées précédemment dont :

- les erreurs issues d'un mauvais découpage lors de l'analyse (*détecteur par canal* qui est une analyse syntaxique erronée de 3 *détecteurs positionnés en parallèle par canal*),
- et les candidats trop généraux (*quantité, grand*).

Ce dernier aspect soulève toutefois plusieurs questions : par exemple, l'un des critères de sélection des termes le plus souvent cité est celui de l'appartenance du terme au domaine étudié. Cependant, dans la mesure où cette recherche vise à repérer l'évolution d'un domaine, il est difficile de circonscrire une liste de termes sur ce critère. Statuer en effet sur l'appartenance d'un terme à un domaine se fait toujours en synchronie⁶⁵ : typiquement, lorsque l'on construit une ressource terminographique, on établit une liste de termes dont on juge qu'ils font partie du domaine au moment de la recherche. En diachronie par contre, la question que l'on devrait se poser – mais à laquelle on ne peut que difficilement répondre *a priori* – serait plutôt : est-ce qu'un candidat a fait partie, fait partie, voire fera partie du domaine dans l'intervalle de temps observé ? En effet, le fait que l'évolution englobe potentiellement des phénomènes de migration, d'apparition de termes ou de disparition implique que le même candidat peut être jugé comme terme du domaine ou non en fonction de la période observée. De fait, le choix posé pour cette recherche est encore une fois d'effectuer une sélection assez large et de ne filtrer que les candidats trop généraux tels que *qualité, grand, partout*, etc. dont on a l'intuition qu'ils n'ont jamais fait, ni ne pourront faire, partie d'un lexique spécialisé. Cette sélection, bien qu'elle soit contrôlée au maximum par l'observation de contextes et de ressources terminographiques complémentaires, reste subjective. Pour la mener à bien, on pourrait envisager de mobiliser des experts du domaine,

⁶⁵ Et pose déjà de nombreuses questions.

comme cela se fait généralement dans les tâches classiques d'extraction de termes. Or, nous n'avons pas mis en place de « tâches expert » dédiées à cette sélection pour deux raisons :

- la première est que ce type de sélection demande beaucoup de temps aux experts qui ne sont que très peu disponibles ; et dans la mesure où ces derniers ont accepté de participer à l'ensemble des autres tâches de cette recherche, il était difficilement envisageable de les solliciter davantage ;
- la seconde raison est liée au type de question qu'auraient dû se poser les experts :
 - o devrait-on demander aux experts de projeter leur jugement dans le temps dès cette étape en leur demandant si les candidats devant eux ont été ou sont encore des termes/concepts de leur domaine ?
 - o ou devrait-on plutôt leur proposer de valider une seule liste sans indication temporelle, au risque que certains candidats soient éliminés du fait de leur caractère obsolète ou trop récent ?

Dans les deux cas, proposer de valider ainsi une liste de candidats nous semblait impliquer de trop grandes difficultés à la fois pratiques (disponibilité des experts) et théoriques (comme nous le verrons au Chapitre 10, la prise en compte du sentiment d'évolution des experts est une question délicate). Pour ces raisons, nous avons choisi de filtrer manuellement la liste sur la base de l'observation des contextes d'apparition des candidats et de critères tels que la productivité du candidat (Bourigault, *et al.*, 2004), ses cooccurrents, la nature de ses actants dans le cas de candidats prédictifs, etc. Néanmoins, lors de chaque tâche proposée aux experts, ces derniers sont invités à supprimer des unités qu'ils jugent ne pas appartenir à leur domaine d'expertise.

Ces différentes étapes de filtrage permettent de réduire considérablement le bruit généré par l'analyse Syntex. Les Tableau 3.4 et Tableau 3.5 illustrent la réduction du nombre de candidats entraînée par ce filtrage pour chacun de nos corpus et pour chacune des catégories lexicales considérées.

	Liste brute avant filtrage	Après programme Perl	Après filtre de fréquence	Après tri manuel final
SNom	20 000	12 332	2394	1117
Nom	3381	2662	1382	708
SV	15 343	4899	591	146
V	1034	1017	585	184
SAdj	1523	816	79	9
Adj	1007	945	571	221
SAdv	92	42	8	0
Adv	473	437	248	10
SNomPr	772	353	49	26
NomPr⁶⁶	1386	823	251	136
Total	45 011	24 326	6 159	2557

Tableau 3.4 – Réduction du nombre de CT après filtrage – corpus TTVS

	Liste brute avant filtrage	Après programme perl	Après filtre de fréquence	Après tri manuel final
SNom	3329	1986	459	317
Nom	1087	863	488	381
SV	2229	807	199	70
V	421	420	207	73
SAdj	104	54	4	0
Adj	299	270	135	44
SAdv	6	6	1	0
Adv	144	136	43	1
SNomPr	178	60	15	0
NomPr	431	192	70	51
Total	8245	4796	1622	937

Tableau 3.5 – Réduction du nombre de CT après filtrage – corpus DORIS

Ces filtres permettent de réduire la liste brute initiale proposée par Syntex de 94,3% pour le corpus TTVS et de 88,6% pour le corpus DORIS et d’obtenir une liste de bonne qualité. La liste de termes ainsi obtenue sert de base au reste de nos explorations.

Ce chapitre nous a permis d’ancrer notre approche en tant qu’approche sur corpus, comparative et outillée pour la diachronie. À partir des données décrites (nos deux corpus,

⁶⁶ NomPr et Les catégories SNomPr (Nom propres), si elles sont supposées rassembler les entités nommées présentes dans le corpus recensent dans les faits les sigles et acronymes présents dans nos deux corpus. Pour cette raison, nous conservons ces classes.

les listes de termes extraits et les outils sélectionnés) nous menons l'identification et la mise en œuvre des quatre indices choisis pour notre approche (cf. §3.1.2.2, p.73).

Nous avons posé dans cette première partie les bases théoriques et méthodologiques au sein desquelles nous ancrons notre recherche linguistique, outillée et diachronique. Nous avons présenté les fondements de l'hypothèse d'un parallèle entre langue et connaissances pour repérer l'évolution mise en œuvre, ainsi que les choix méthodologiques qui président au repérage de l'évolution en corpus que nous proposons.

Nous abordons maintenant la description de quatre indices linguistiques pour le repérage. Nous l'avons dit, les indices présentés ont été choisis sur la base d'allers-retours entre notre hypothèse et les données à l'aide d'une analyse outillée minutieuse dans nos deux corpus. Dans la partie suivante, nous décrivons chacun des indices et justifions leur pertinence pour une analyse diachronique. Nous détaillons la manière dont ils sont mis en œuvre au sein de l'approche comparative et outillée que nous défendons. La pertinence de chacun des indices est mesurée et permet de caractériser les différentes facettes de l'évolution des connaissances en diachronie courte.

Cette deuxième partie est divisée en cinq chapitres. Les quatre premiers sont dédiés à chacun des indices sélectionnés (respectivement les empreintes de fréquence, les contextes riches en connaissances évolutives, les variantes de termes et les dépendances syntaxiques). Le cinquième propose un bilan des observations mises au jour grâce à ces indices, ainsi qu'une typologie de l'évolution des connaissances du domaine spatial en diachronie courte.

Deuxième partie

Exploration outillée de quatre indices
linguistiques

Chapitre 4 Empreintes de fréquence

SOMMAIRE DU CHAPITRE

4.1	EMPREINTES DE FRÉQUENCE ET CORPUS DIACHRONIQUES.....	106
4.1.1	« Empreintes de fréquence » : définition.....	106
4.1.2	Néologie et obsolescence.....	107
4.1.3	Formes chrono-homogènes.....	109
4.2	MÉTHODOLOGIE POUR LA MISE AU JOUR D'EMPREINTES.....	110
4.2.1	Sélection du seuil de fréquence.....	110
4.2.2	Définition d'empreintes de fréquence : mise en œuvre du χ^2	111
4.3	OBSERVATION DES EMPREINTES DE FRÉQUENCE EN CORPUS.....	115
4.3.1	Apparition et disparition de termes.....	115
4.3.1.1	Corpus TTVS.....	115
4.3.1.2	Corpus DORIS.....	124
4.3.1.3	Synthèse.....	130
4.3.2	Termes chrono-homogènes.....	132
4.3.2.1	Repérage de thématiques.....	132
4.3.2.2	Interprétations des experts.....	134
4.3.2.3	Synthèse.....	142
4.4	BILAN.....	144

Dans ce chapitre, nous présentons un premier indice pour repérer l'évolution des connaissances en corpus : les empreintes de fréquence. Lorsque l'on travaille sur corpus, les données chiffrées s'imposent et sont d'ailleurs le plus souvent la première étape d'exploration mise en œuvre dans les textes. En linguistique de corpus, le calcul des fréquences est au cœur du traitement et de l'analyse des données, que ce soit en synchronie ou en diachronie et les données chiffrées sont de plus en plus faciles et rapides à obtenir grâce aux outils actuels. L'analyse des fréquences permet de mettre au jour les variations de certains traits linguistiques, des tendances, des phénomènes saillants, qui, en diachronie, peuvent être associés à une évolution des connaissances.

Dans la première section de ce chapitre, nous définissons les empreintes de fréquence et illustrons leur mise en œuvre en diachronie. Dans la deuxième section, nous décrivons notre démarche de calcul et de sélection des empreintes sur la base du χ^2 . Nous terminons par une description détaillée des phénomènes d'évolution observables dans nos corpus à l'aide de ces empreintes et montrons la pertinence de cette approche pour la diachronie.

4.1 Empreintes de fréquence et corpus diachroniques

4.1.1 « Empreintes de fréquence » : définition

Le terme « empreintes de fréquence », introduit par Ahmad *et al.* (2002)⁶⁷, renvoie à la possibilité de dessiner la courbe de fréquence d'un terme à travers les différents sous-corpus composant un corpus diachronique. L'hypothèse est de considérer que la courbe de fréquence d'un terme donné dans le corpus reflète l'évolution du terme/concept dans le domaine. C'est ce qu'explique Losee (1995 : 269) dans l'extrait suivant :

« We assume that the change in interest in concepts may be tracked back by examining the presence or absence of terms in documents. [...] [C]oncepts and the associated terms develop and become popular over a period of years. [...] As concepts grow in disciplinary importance, we expect terms associated with concepts to increase in frequency. Intra-disciplinary growth begins with a growth period that then levels off after several years. When a concept becomes discredited or is replaced, or has been exhaustively studied, a decrease in use will occur. »

Comme l'illustre cette citation, plusieurs types d'empreintes peuvent être définis en fonction de la courbe de fréquence et parmi eux, quatre principaux⁶⁸ :

- l'apparition d'un terme dans les sous-corpus les plus récents,
- la disparition d'un terme dans les sous-corpus les plus récents,
- la fréquence croissante du terme dans l'ensemble des corpus,
- et la décroissance du terme dans l'ensemble des corpus.

Ces quatre empreintes peuvent être classées en deux groupes sur la base de la distinction possible entre rupture et continuité.

Les deux premières empreintes peuvent être associées à la notion de « rupture » dans la mesure où elles passent par une fréquence nulle dans au moins l'un des sous-corpus qui constituent le corpus comparable (§4.1.2). La première d'entre elles (l'apparition d'un terme)

⁶⁷ « Empreinte de fréquence » est notre traduction de « Frequency Signature » proposée à l'origine par Ahmad *et al.*

⁶⁸ Ces quatre empreintes sont possibles dans le cas où un corpus comparable contient 2 sous-corpus. Dans le cas du corpus TTVS qui contient 3 sous-corpus, 2 autres empreintes sont possibles : croissance/décroissance et décroissance/croissance.

a fait l'objet d'attentions particulières de la part des néologues dans la mesure où ce phénomène est susceptible de refléter l'apparition d'un néologisme dans le domaine. À l'inverse, la disparition d'un terme du corpus peut révéler la disparition d'un terme/concept obsolète dans le domaine.

Les deux dernières empreintes reposent quant à elles sur des schémas continus : les termes considérés n'apparaissent ni ne disparaissent spécifiquement, ce sont leurs variations de fréquence dans l'ensemble des sous-corpus qui sont observées et qui peuvent être associées à la croissance/décroissance des termes/concepts dans le domaine. Ce type de courbe est particulièrement intéressant pour tenter de repérer des termes appelés « chrono-homogènes » (Habert, *et al.*, 1997)(§ 4.1.3).

4.1.2 Néologie et obsolescence

Comme le souligne Humbley (2006 : 92-93), définir ce qu'est un néologisme n'est pas chose facile, et lorsqu'il s'agit d'en donner une définition « plus pratique que théorique, lexicographique plutôt que lexicologique », on fait appel « au dispositif *ad hoc* du corpus d'exclusion » dans lequel « le néologisme est un mot relevé dans un texte mais qui ne se trouve pas dans un corpus de dictionnaires supposé représenter l'état [...] de la langue et du registre concernés ».

Cette définition au caractère méthodologique se traduit en corpus par la prise en compte d'une empreinte de fréquence : un terme absent d'un corpus ancien, dit de référence, mais présent dans un corpus récent peut être considéré comme un néologisme ou, plus précisément, comme un néologisme potentiel.

Cette définition offre des perspectives intéressantes pour le repérage automatisé des néologismes, automatisation qui « permet d'alléger considérablement la tâche du linguiste » (Mathieu, *et al.*, 1998 : 207). Rappelons que bon nombre de travaux ont vu le jour sur cette question (Cabré, *et al.*, 2003 ; Janssen, 2008 ; Mathieu, 1998 ; Mathieu, *et al.*, 1998 ; Ollinger & Valette, 2008 ; Roche & Bowker, 1999 ; Sader Feghali, 2005, etc.), mais que ce n'est que récemment que des propositions ont été faites sur des corpus comparables seuls (Drouin, *et al.*, 2006 ; Paquin, 2007) (§1.2.3.2, p.28).

À l'inverse, l'empreinte de fréquence d'un terme peut indiquer que celui-ci est devenu obsolète. Cet aspect de la disparition des termes est très peu abordé en terminologie, mais l'on trouve parfois le terme de *mort lexicale* pour le nommer, par exemple chez Grzega (2002), ou plus récemment le terme de *nécrologie* chez Drouin et Dury (à paraître). Ces derniers mettent en place une des rares études sur cet aspect, étude basée sur un repérage semi-automatique des « nécrologismes » à partir de la comparaison de corpus, sur le modèle du repérage des néologismes.

Dans cette recherche, nous choisissons le terme *obsolescence* plutôt que *nécrologie*, qui sous-tend l'idée de la disparition totale d'un terme/concept (de sa « mort » donc) alors qu'il nous semble plutôt qu'une des particularités de la disparition de termes/concepts par rapport à leur apparition est qu'ils peuvent toujours rester dans la mémoire d'experts, qu'ils sont susceptibles de réapparaître dans la terminologie et surtout qu'il en reste toujours des traces dans des textes anciens et archives consultables. Pour cette raison, et ceci est particulièrement notable en diachronie courte, la disparition des termes relève plus de leur obsolescence que de leur « mort ».

Soulignons que la majorité de ces études s'accordent sur le fait que le critère de fréquence seul n'est pas un indice suffisant pour décider de la néologicité – ou de l'obsolescence – d'un terme repéré. Beaucoup de ces systèmes font de fait souvent appel à une série de filtres supplémentaires tels que des filtres morphologiques (en fonction du mode de formation d'un candidat) ou statistiques, pour affiner le repérage. Mais malgré tout le soin apporté à la sélection des candidats néologismes, « il reste [...] [au linguiste] un travail important à faire : établir si un néologisme potentiel est ou non un vrai néologisme » (Mathieu, *et al.*, 1998 : 207). La validation et l'interprétation de ces candidats est en effet une étape généralement incontournable pour construire une liste de néologismes ou termes/concepts obsolètes acceptables. Et c'est sur cet aspect que nous souhaitons insister dans ce chapitre : si les termes qui apparaissent uniquement dans les corpus les plus récents ne sont pas forcément des néologismes, quelles autres interprétations peut-on leur associer ? De la même manière, si des termes disparaissent des corpus les plus récents, sont-ils nécessairement obsolètes ?

4.1.3 Formes chrono-homogènes

Le second type d'observations possibles à partir de la fréquence repose sur une dynamique de « continuité » : on ne cherche pas spécifiquement des termes qui apparaissent/disparaissent dans le corpus, mais on observe la courbe de fréquence des termes dans l'ensemble des corpus. Ce type d'hypothèses est fréquent en diachronie et permet d'associer des courbes de fréquences dans un corpus à l'évolution d'un terme/concept dans le domaine. Ainsi par exemple, Dury (2007) argumente une analyse de la « terminologisation » du terme *carbon neutral* dans le domaine de l'écologie à l'aide de l'augmentation de fréquence de ce terme dans chacun des sous-corpus spécialisés observés. Ou encore, Ahmad et Musacchio (2004 : 1568) montrent l'évolution des usages du terme *neutroni* qui diminuent dans le temps, alors que ceux du terme *quark* augmentent révélant ainsi le changement d'intérêt sur la question des particules élémentaires en physique. Enfin, Schierz (2007 : 133-134) illustre le potentiel de ces courbes de fréquence pour comparer l'évolution de termes dans différents types de documents et montre que des pics de croissance de termes sont repérables dans des corpus de textes universitaires et académiques bien avant que ces termes n'apparaissent dans des corpus de brevets.

L'intérêt de ces empreintes de fréquence ne réside pas seulement dans leur utilisation pour l'observation de termes isolés, mais également pour contraster, comparer et regrouper certains termes qui présentent des empreintes de fréquence similaires dans le temps. On retrouve cette idée chez Habert *et al.* (1997 : 209-210), qui parlent de formes « chrono-homogènes » :

« Les schémas d'évolution établis pour chacune des unités font apparaître des ensembles d'unités qui ont tendance à évoluer de conserve au fil des périodes : les formes *chrono-homogènes*.

En fait, l'idée qui sous-tend cette approche est la suivante : pour des formes fréquentes dans le corpus, le fait que plusieurs formes évoluent de manière proportionnelle tout au long des périodes ne peut être mis au compte du hasard. Il faut donc, dans chaque cas, déterminer la cause profonde qui est à l'origine de ces regroupements. Selon les cas, on trouvera des groupements liés à une thématique, à une actualité, etc. »

L'avantage de ces regroupements est qu'en plus de l'observation de termes isolés, ils permettent de faire émerger des thématiques pour observer l'évolution des connaissances. Comme nous le verrons (§4.3.2), ceci a au moins deux intérêts :

- l'émergence de thématiques d'évolution permet de compléter les observations classiques de l'évolution des terminologies en leur faisant dépasser le niveau du terme/concept isolé souvent privilégié dans les descriptions terminologiques ;
- le fait que certains termes/concepts évoluent « en masse » permet d'aider l'analyste dans son interprétation de l'évolution, de l'affiner et de donner plus de poids à son argumentation.

Mais avant de développer ces deux points, nous expliquons dans la section suivante la méthode choisie dans cette étude afin de mettre au jour des empreintes de fréquence significatives pour observer l'évolution.

4.2 Méthodologie pour la mise au jour d'empreintes

4.2.1 Sélection du seuil de fréquence

Puisque l'un des objectifs de ce chapitre est de voir dans quelle mesure on peut interpréter des phénomènes d'évolution des connaissances à partir de variations de fréquence en corpus, il est nécessaire de s'assurer que les variations observées sont les plus significatives possibles. Ceci est d'autant plus important lorsque l'analyse porte sur des corpus de petite taille (comme c'est souvent le cas en langue de spécialité et en diachronie), qui rendent l'indice de fréquence difficile à manipuler. En effet, il peut être difficile de définir un seuil de croissance ou de décroissance significatif d'une évolution : à partir de quel degré de croissance/décroissance peut-on associer une variation de fréquence significative à une évolution dans le domaine ? La fréquence doit-elle doubler, tripler ? Doit-on considérer le même seuil pour les basses fréquences que pour les hautes fréquences ?

Pour répondre à ces questions, il est généralement possible de se concentrer sur l'observation des termes les plus fréquents dans les corpus en appliquant un seuil minimal de fréquence arbitraire (par exemple les termes de plus de 100 occurrences). Mais dans le cas de petits corpus comme ceux choisis pour cette recherche, un tel seuil est difficile à appliquer dans la

mesure où peu de termes sont très fréquents (notamment dans le corpus DORIS qui compte au total à peine plus de 36 000 occurrences). De plus, dans le cas du corpus TTVS, les sous-corpus choisis sont très déséquilibrés (§3.2.2.1) et il est nécessaire de mettre en place un seuil fiable pour s'assurer que les termes en croissance ou les termes apparus ne sont pas liés à la seule augmentation de taille des sous-corpus, mais bien à une évolution du domaine.

Soulignons également que le choix d'un filtre qui garantisse au maximum la significativité des termes traités, permet de réduire et filtrer la liste des résultats avant de les présenter aux experts. En effet, il est important de sélectionner des résultats peu nombreux mais significatifs pour les experts plutôt que de longues listes très bruitées qui, en plus de demander beaucoup de temps de traitement, risquent de les faire douter de l'intérêt de la démarche linguistique (cf. §10.2.2.3, p.307).

Ajoutons enfin que, si l'on souhaite reproduire ce type d'analyse sur d'autres corpus, il est nécessaire de savoir sélectionner au mieux le seuil de fréquence à appliquer.

Pour répondre à ces différentes contraintes et mettre au jour des empreintes de fréquence significatives et reproductibles, une solution est le recours à un calcul statistique tel que le test du χ^2 (Chi-Deux ou Khi-Deux) :

« [le test du χ^2] sert à apprécier en probabilité l'écart constaté entre une observation et un modèle théorique, quel que soit le nombre des variables. »
(Muller, 1968 : 116)

Ce calcul a été choisi pour appliquer l'indice de fréquence dans notre recherche⁶⁹.

4.2.2 Définition d'empreintes de fréquence : mise en œuvre du χ^2

Le calcul du χ^2 , et plus précisément pour cette étude le χ^2 d'ajustement (Howell, 2008 : 142-144), implique une comparaison entre des fréquences observées et des fréquences attendues ou théoriques :

⁶⁹ Nous avons été aidée pour cela par Nathan Ménard et Patrick Drouin, respectivement Professeur de statistiques linguistiques et Professeur en terminologie et traduction à l'Université de Montréal.

« Les fréquences observées [...] sont les fréquences effectivement observées dans les données [...]. Les fréquences attendues sont les fréquences que nous nous attendons à voir *si l'hypothèse nulle est vraie*. » (Howell, 2008 : 143)

L'hypothèse nulle testée ici est que les variations de fréquences sont dues aux variations de taille des sous-corpus (et non pas à une évolution du domaine). Le χ^2 permet donc de comparer les fréquences observées dans chacun des sous-corpus et les fréquences attendues que l'on devrait normalement observer dans ces sous-corpus si la variation était simplement due à leur variation de taille. Pour ce faire, on applique la formule suivante (Figure 4.1) :

$$\chi^2 = \sum \frac{(O - A)^2}{A}$$

où :

O = fréquence observée dans le corpus

A = fréquence attendue ou théorique

Figure 4.1 – Calcul du χ^2 (Howell, 2008 : 143)

Pour mener ce calcul sur nos corpus, nous nous basons sur un script Perl disponible en ligne et proposé par Karlgren⁷⁰. En sortie, le calcul du χ^2 est présenté sous la forme de tables de contingence (Tableau 4.1) pour chacun des termes traités.

		Fréquences de <i>polissage</i>	Fréquences des autres termes dans les sous-corpus
Fréquences observées	TTVS1994	2	8445
	TTVS1998	28	14853
	TTVS2002	53	20452
Fréquences attendues	TTVS1994	15,994	8431,005
	TTVS1998	28,178	14852,822
	TTVS2002	38,83	20466,172
χ^2		17,452	

Tableau 4.1 – Table de contingence – Calcul du χ^2 du terme *polissage* (TTVS)

⁷⁰ <http://www.sics.se/~jussi/Verktyg/khi2.perl> (consulté le 25 octobre 2007).

Une fois la valeur du χ^2 connue, la seconde étape concerne son interprétation. En effet, la valeur de χ^2 obtenue doit être comparée à la distribution χ^2 en table, consultable dans les ouvrages de statistiques tels que ceux de Howell (2008 : 144) ou Muller (1968 : 179) utilisés ici et dont on présente un extrait dans le Tableau 4.2.

Cette table indique une probabilité P en fonction des degrés de liberté pris en compte dans le calcul de χ^2 (noté aussi *ddl*, Howell, 2008 : 146). Dans notre cas, 2 degrés de liberté sont pris en compte pour le corpus TTVS et 1 degré de liberté dans le corpus DORIS⁷¹. Plus la valeur de P est petite, plus le score de χ^2 est significatif. Si l'on reprend l'exemple de *polissage* dans le TTVS pour lequel $\chi^2 = 17,452$ et $P < 0,001$, ceci signifie qu'il y a moins d'1 chance sur 1000 que l'hypothèse nulle soit validée, c'est-à-dire que la variation observée soit simplement due à la variation de taille des sous-corpus. Dans le cas de *polissage*, la variation observée peut donc être associée à une autre cause, dont nous faisons l'hypothèse dans cette recherche qu'il s'agit d'une évolution du domaine.

χ^2 \ P	0,90	0,70	0,50	0,30	0,10	0,05	0,02	0,01	0,001
1	0,016	0,148	0,455	1,074	2,706	3,841	5,024	6,635	10,827
2	0,211	0,713	1,386	2,408	4,605	5,991	7,824	9,210	13,815

Tableau 4.2 – Extrait du tableau de distribution du χ^2 (Muller, 1968 : 179)

Dans cette recherche, nous ne retiendrons que les cas de χ^2 où $P = 0,01$ au maximum, c'est-à-dire les cas où la probabilité que la variation obtenue soit liée à la taille du corpus est d'une chance sur 100 au maximum (soit $\chi^2 \geq 9,210$ pour le TTVS et $\chi^2 \geq 6,635$ pour DORIS).

Une fois ce calcul appliqué à l'ensemble des corpus, une liste de termes dont les variations de fréquence sont statistiquement significatives est obtenue pour servir de base à l'analyse.

Sur la base de cette liste, il est possible de dégager des empreintes de fréquence, c'est-à-dire les courbes de variation de ces termes. Pour ce faire, nous proposons de nous baser sur la notion d'*écart*. L'écart est la différence entre la fréquence observée et la fréquence attendue calculée par le χ^2 . Par exemple, dans le cas de *polissage*, les écarts obtenus sont les suivants :

⁷¹ Le nombre de degrés de liberté est calculé de la manière suivante : $ddl = (L-1)(C-1)$ où L = le nombre de lignes de la table de contingence (soit 3 dans le cas du TTVS) et C = le nombre de colonnes de la table de contingence (soit 2 dans le cas du TTVS).

<i>polissage</i>	Fréquence observée	Fréquence attendue	Écart
TTVS1994	2	15,994	-13,995
TTVS1998	28	28,178	-0,178
TTVS2002	53	38,83	14,173

Tableau 4.3 – Écarts obtenus – exemple de *polissage* (TTVS)

Dans ce cas, les écarts sont négatifs pour les deux premiers sous-corpus et l'écart devient positif dans le TTVS2002. Autrement dit, la fréquence observée du terme *polissage* est moins importante qu'attendue dans les deux premiers sous-corpus et plus importante qu'attendue dans le dernier. Ce terme est donc en croissance à partir de 2002. Sur ce modèle, plusieurs empreintes peuvent être définies et observées :

1. **Pas d'évolution** : les variations de fréquences ne sont pas significatives entre les sous-corpus, c'est-à-dire que le χ^2 obtenu correspond à une probabilité trop forte d'hypothèse nulle ;
2. **Croissance** : la fréquence observée est plus faible qu'attendue dans les sous-corpus les plus anciens et plus forte qu'attendue dans les sous-corpus les plus récents. Dans cette catégorie peuvent être classés les cas de néologismes potentiels ;
3. **Décroissance** : à l'inverse, la fréquence observée est plus forte qu'attendue dans les sous-corpus les plus anciens et est plus faible qu'attendue dans les sous-corpus les plus récents. De la même manière, les cas de termes/concepts obsolètes potentiels peuvent être classés dans cette catégorie ;
4. **Pic de croissance** : cette empreinte n'est possible que dans le cas où au moins trois sous-corpus sont observés, comme dans le cas du TTVS. Dans ce corpus, certains termes peuvent ainsi connaître une fréquence observée plus forte qu'attendue uniquement dans le sous-corpus intermédiaire (ici le TTVS1998) ;
5. **Pic de décroissance** : à l'inverse, certains termes peuvent ainsi avoir une fréquence observée plus faible qu'attendue uniquement dans le sous-corpus intermédiaire.

Le Tableau 4.4 illustre chacune de ces empreintes pour le corpus TTVS.

	Termes	χ^2	Écart TTVS1994	Écart TTVS1998	Écart TTVS2002
1. Pas d'évolution	<i>orbite géostationnaire</i>	0,081	0,641	0,203	0,437
2. Croissance	<i>polissage</i>	17,452	-13,995	-0,178	14,173
3. Décroissance	<i>refroidir</i>	30,687	20,693	-14,252	-6,441
4. Pic de croissance	<i>microns</i>	16,752	-13,115	15,088	-1,973
5. Pic de décroissance	<i>système optique</i>	10,134	3,219	-8,185	4,966

Tableau 4.4 – Exemples pour chaque empreinte de fréquence

Dans la suite de ce chapitre, nous analysons dans un premier temps les termes apparus/disparus (qui correspondent aux empreintes de rupture) pour chacun des deux corpus. Dans la mesure où ils correspondent à des termes/concepts potentiellement néologiques/obsolètes, l'interprétation de l'évolution auprès d'experts permettra d'observer dans quelles mesures le filtre de fréquence permet de repérer ce genre de cas en diachronie courte. Dans un second temps, ce sont les empreintes continues qui feront l'objet d'une analyse. Ces empreintes seront appliquées pour le repérage de formes chrono-homogènes, qui permet cette fois-ci non plus de traiter des termes isolés, mais plutôt de les regrouper en séries de termes évolutifs.

4.3 Observation des empreintes de fréquence en corpus

4.3.1 Apparition et disparition de termes

4.3.1.1 Corpus TTVS

Dans la mesure où le corpus TTVS est composé de trois sous-corpus, les termes absents du premier sous-corpus mais présents dans les deux plus récents **ou** les termes absents dans les deux premiers sous-corpus mais présents dans le plus récent seront considérés comme termes apparus. Par exemple :

Forme	Fréquence dans les sous-corpus		
	TTVS1994	TTVS1998	TTVS2002
<i>APS</i>	0	2	70
<i>arêtes</i>	0	9	30
<i> fibre dopée</i>	0	12	15
<i>facteur de remplissage</i>	0	0	13
<i>adhérence</i>	0	0	12

Tableau 4.5 – Exemples de termes apparus (TTVS)

Les termes présents dans le premier sous-corpus mais absents dans les deux plus récents **ou** les termes présents dans les deux premiers sous-corpus mais absents dans le plus récent seront considérés comme termes disparus. Par exemple :

Forme	Fréquence dans les sous-corpus		
	TTVS1994	TTVS1998	TTVS2002
<i>Modèle Instrumental</i>	11	0	0
<i>instruments imageurs</i>	3	5	0
<i>Modèle Radiométrique Instrumental</i>	7	0	0
<i>synthèse d'ouverture optique</i>	5	2	0

Tableau 4.6 – Exemples de termes disparus (TTVS)

Au total dans ce corpus, avant que l'on applique un filtre sur la base du χ^2 , 278 termes apparaissent entre la première et la dernière édition et 34 disparaissent. Soulignons que ces 312 (278+34) termes apparus/disparus représentent environ 12% de l'ensemble des termes retenus pour cette analyse (2557 au total, cf. Tableau 3.4, p.101). On observe également une plus grande proportion de termes qui apparaissent que de termes qui disparaissent dans le TTVS (89% contre 11% qui disparaissent). Bien que ce dernier aspect puisse être dû en grande part à l'augmentation de taille des sous-corpus dans le temps, la même observation reste valable dans la version équilibrée du corpus TTVS, utilisée notamment dans (Picton, 2007) et décrite dans le chapitre précédent. En d'autres termes, même dans une version équilibrée du corpus, on observe toujours plus de termes qui apparaissent que de termes qui disparaissent. On peut donc dire que, globalement, la terminologie du domaine de l'optique telle que la reflète le TTVS est en croissance entre 1994 et 2002.

Pour mener des observations détaillées en corpus, seuls les termes dont le χ^2 indique que la variation observée est statistiquement significative (selon le seuil choisi pour cette recherche : $\chi^2 \geq 9,210$) sont sélectionnés. Ajoutons cependant qu'un filtre de fréquence a été ajouté⁷² aux termes retenus par le χ^2 : seuls les termes dont la fréquence minimale est de 30 dans l'ensemble du corpus TTVS et dont le χ^2 est significatif ont été retenus. En effet, si le χ^2 peut être appliqué à tous les termes, il nous a été suggéré de ne conserver que les termes de

⁷² Sur une suggestion de Nathan Ménard.

fréquence supérieure ou égale à 30 pour s'assurer de la pertinence du χ^2 dans le TTVS. Ce filtre statistique réduit considérablement le nombre de termes pris en compte puisque seuls 7 termes apparus sont significatifs selon ce critère (Tableau 4.7) alors qu'aucun des termes disparus ne l'est. Cette réduction importante du nombre de candidats ne signifie cependant pas qu'aucun des 271 autres termes apparus n'est pertinent par rapport à l'évolution du domaine, mais plutôt qu'ils ont moins de chance de pouvoir y être significativement associés.

Dans la mesure où aucun des termes disparus ne compte plus de 12 occurrences, aucun n'est significatif sur la base du filtre défini. Cependant, mais sous réserve de représentativité, nous décrivons les 5 termes disparus les plus fréquents (Tableau 4.8) afin d'observer des tendances potentielles de changement, ainsi que le sentiment d'évolution des experts sur ces termes.

Forme	TTVS1994	TTVS1998	TTVS2002	Fréquence totale	χ^2
<i>APS</i>	0	2	70	72	72,187
<i>HRG</i>	0	29	29	58	15,728
<i>arêtes</i>	0	9	30	39	16,463
<i>cavité étendue</i>	0	16	23	39	9,349
<i>colle</i>	0	1	34	35	35,717
<i>transition</i>	0	8	22	30	10,779
<i>cœur</i>	0	3	27	30	22,845

Tableau 4.7 – 7 « Candidats néologismes » significatifs ($\chi^2 \geq 9,210$ et $\text{freq} \geq 30$) –TTVS

Forme	TTVS1994	TTVS1998	TTVS2002	Fréquence totale	χ^2
<i>Modèle Instrumental</i>	11	0	0	11	45,938
<i>instrument imageur</i>	3	5	0	8	7,031
<i>Modèle Radiométrique Instrumental</i>	7	0	0	7	29,245
<i>synthèse d'ouverture optique</i>	4	3	0	7	8,625
<i>LDEF</i>	4	3	0	7	8,625

Tableau 4.8 – 5 « Candidats obsolètes » (5 termes les plus fréquents disparus) –TTVS

Les observations suivantes sont donc basées respectivement sur ces 7 et 5 termes présentés aux experts.

A. Apparition de termes

Comme nous l'avons vu, chacun des 7 termes du Tableau 4.7 peut être considéré comme un néologisme potentiel. Sur la base de cet échantillon, les quatre experts en optique qui ont accepté de participer à notre étude ont été interrogés sur leur sentiment face à l'évolution de ces termes/concepts. La tâche suivante leur a été proposée : *Selon vous, les termes présentés dans cette liste sont-ils nouveaux dans le domaine de l'optique et de l'optoélectronique spatiales ? Sinon, pouvez-vous les lier à une évolution du domaine ?*

À partir de cette tâche, les experts ont unanimement jugé qu'un seul de ces termes pouvait effectivement être considéré comme un néologisme, le terme *APS*. Pour tenter de mieux comprendre l'apparition des autres termes dans le corpus, nous rapportons ici certains éléments révélés lors des entretiens qui reflètent le sentiment des experts sur l'évolution de chacun de ces termes/concepts d'optique et optoélectronique. Ces éléments sont complétés par des informations extraites du site du Cnes⁷³.

Le premier terme, *HRG*, renvoie à un type d'instruments d'observation embarqués sur le satellite Spot 5 : les instruments à Haute Résolution Géométrique. Bien que ce terme n'apparaisse qu'à partir de 1998, il s'agit d'instruments liés au projet Spot 5, développé dès 1994 en continuité de la filière des satellites Spot démarrée en 1986⁷⁴ (Cnes-Presse, 2003). De fait, lorsque ce terme/concept apparaît dans le TTVS en 1998, ce type d'instruments est connu et élaboré par des experts Cnes depuis plusieurs années. Il s'agit donc d'instruments récents en 1998, mais qui auraient pu éventuellement être mentionnés dans l'édition de 1994. Ce terme/concept apparaît donc un peu trop familier des experts en 1998 pour qu'ils le jugent nouveau. Notons cependant que ces instruments sont bien liés à des progrès importants des techniques de détection sur Spot 5, en comparaison avec les quatre générations précédentes de satellites Spot :

« Des progrès spectaculaires améliorant l'acquisition des images ont été réalisés dans un grand nombre de technologies, par exemple :

⁷³ <http://www.cnes.fr>

⁷⁴ Spot 5 est la troisième génération d'instruments Spot après Spot 1 à 3 pour la première génération et Spot 4 pour la deuxième.

- pour la détection : de nouveaux détecteurs en ligne 12 000 points : La lumière réfléchiée par le paysage est captée par les HRG, instruments optiques puissants et très imposants. [...] Ces lignes de détecteurs sont constituées d'un seul élément (appelé barrette) au lieu de quatre sur les satellites précédents. Ce gain de place a permis de simplifier le plan focal et d'ajouter une bande spectrale. » (Cnes-Presse, 2003 : 2).

Le deuxième terme, *arêtes*, renvoie à un concept ancien et fondamental en optique. D'emblée, les experts interrogés l'ont donc exclu de la liste des candidats néologismes. Pourtant, l'apparition de ce terme dans le corpus à partir de 1998 peut être liée à une évolution du domaine. En poursuivant l'exploration en corpus et avec les experts, il est apparu que ce terme est lié dans le cours à la question des *parasites* et de leur traitement. Un de nos experts a insisté sur cet aspect en expliquant qu'une arête est un bord tranchant sur lequel on ne peut plus définir la notion de surface. Or, on peut ne voir la lumière que lorsqu'elle apparaît sur une arête (par exemple dans le cas des objets transparents qui n'apparaissent que par leur contour : la lumière se casse sur ces bords tranchants). Le concept d'*arêtes* intervient donc nécessairement dans la définition de la lumière parasite. Et de nombreux progrès ont eu lieu quant à la question des parasites et de leur traitement au Cnes (cf. §7.3.4, p.243). De fait, bien qu'il ne s'agisse pas d'un terme/concept nouveau dans le domaine de l'optique, c'est un concept nécessaire pour décrire des progrès importants dans le domaine. Autrement dit, cet exemple montre que les évolutions scientifiques et techniques peuvent impliquer le recours à des termes/concepts anciens pour décrire ces développements. Dans ce cas, les termes/concepts apparus ne sont donc pas des termes récents ou nouveaux, mais des termes/concepts « rappelés » dans les textes dans la description d'évolutions auxquelles ils participent.

Le même genre de phénomène peut être décrit avec l'apparition du terme *cœur* qui est un terme/concept ancien dans le domaine de l'optique⁷⁵. Cependant, en observant ses contextes d'apparition d'un peu plus près, on s'aperçoit rapidement que lorsqu'on le rencontre en 2002, il est essentiellement associé au terme/concept *fibres optiques* (*cœur de fibre*, *cœur de la fibre*). Les fibres optiques sont les technologies les plus récentes dans la technologie des *lasers* et font

⁷⁵ Il s'agit également d'un terme assez général que les experts ont cependant choisi de traiter dans la mesure où il est fréquemment utilisé en optique, généralement dans des syntagmes tels que *cœur de gaine*, *cœur optique*, etc.

l'objet de grands développements dans le spatial depuis les années 90. De fait, encore une fois, ce terme fondamental et ancien en optique apparaît dans le corpus dans la mesure où il accompagne la description du développement d'une technologie récente dans le domaine.

Le quatrième terme, *cavité étendue*, est un syntagme rarement utilisé seul dans le domaine. Il est généralement associé aux termes *diode* et *laser*. Les diodes et lasers ne sont encore une fois pas des éléments nouveaux en optique, mais le fait que la cavité du laser soit étendue est une propriété récente dans le spatial, liée à la recherche de meilleures performances pour des projets tels que Pharaos, une horloge atomique dont l'étude a commencé au début des années 1990 (Cnes, mars 2007). De plus, bien que les lasers soient développés en optique depuis les années 60, leur arrivée dans le spatial ne date réellement que des années 80, soit relativement récemment au vu de la période observée dans cette recherche (1994-2002). De manière générale, les lasers sont des éléments essentiels en optique spatiale, qui suscitent de nombreux développements dans le domaine, développements que différents exemples présentés dans ce travail viendront souligner.

Le cinquième terme de la liste, *colle*, n'est pas un terme/concept propre au domaine de l'optique. Cependant des colles spécifiques sont utilisées en optique spatiale pour lier les miroirs et les lentilles sur différentes structures. Il s'agit donc d'un élément essentiel dans l'élaboration de systèmes optiques dans la mesure où les colles utilisées doivent garantir une fixation maximale dans l'environnement spatial. Le problème du choix et de l'utilisation des colles s'est toujours posé aux opticiens, mais il s'est accentué avec la progression des applications spatiales et de certains problèmes qui leur sont propres : une gamme de températures plus grande, des problèmes d'irradiation dans le vide, etc. De ce point de vue, les experts sont conscients de l'évolution rapides des colles et des contraintes auxquelles elles doivent répondre. Selon un des experts rencontrés sur cette question, les colles sont réellement au centre de progrès et d'évolutions en optique spatiale. Cependant, ce point de vue n'est pas partagé par l'ensemble des experts interrogés. Pour la majorité en effet, il reste difficile d'expliquer avec certitude pourquoi cette question n'apparaît précisément et massivement que dans l'édition 2002, alors qu'elle aurait pu être abordée dans les éditions précédentes du cours. Malgré les progrès vraisemblablement apportés à l'élaboration des

colles utilisées en optique, ces experts ne peuvent lier l'apparition de ce terme/concept dans le TTVS 2002 avec un projet spatial spécifique ou un autre événement qui puisse justifier l'absence de ce terme/concept dans les éditions 1994 et 1998. Il est donc probable que l'apparition de ce terme dans l'édition 2002 soit plutôt due à un enrichissement du cours qu'à une évolution des connaissances du domaine. Dans ce cas, il s'agit alors certainement d'une thématique ajoutée dans le cours pour le compléter en tant que référence, et non pas pour le mettre à jour par rapport aux évolutions du domaine.

Enfin, le terme *transition* est un peu plus difficile à traiter dans la mesure où il s'agit d'un terme général. Cependant, à travers ses contextes d'utilisation, il peut être lié à deux autres termes/concepts dans le corpus :

- le terme/concept *transition optique*, utilisé essentiellement dans les sections abordant la question des lasers à fibre dopée et les lasers pompés par diode,
- le terme/concept *transition vitreuse*, utilisé à propos des colles pour parler des évolutions structurelles subies par les colles à certaines températures.

On retrouve donc ici deux thématiques mentionnées plus haut et liées à une évolution du domaine : les lasers en développement constant en optique spatiale depuis les années 80 et les colles qui sont nouvellement abordées dans le cours TTVS (sous réserve qu'il s'agisse bien d'une évolution et non pas d'un enrichissement du cours comme le pensent certains experts). Ce terme/concept n'est pas nouveau, modifié ou récent dans le domaine, mais accompagne plutôt la description d'autres évolutions avec lesquelles il peut être mis en lien.

Le détail de ces quelques termes apparus dans le domaine permet de souligner d'emblée la diversité des aspects d'évolution qu'une empreinte de fréquence telle que l'apparition de termes peut, d'après les experts, indiquer. L'observation de cette courte liste de 7 termes/concepts permet déjà d'indiquer que l'apparition d'un terme dans le domaine peut être expliquée par :

- l'apparition d'un néologisme (*APS*),
- la récence d'un terme/concept dans le domaine (*HRG*),
- la mise en place de projets spatiaux (*HRG*),

- le rappel de concepts bien établis dans le domaine pour la description d'autres évolutions (*arêtes, cœur*),
- l'enrichissement du corpus (*colle*).

Il s'agit maintenant d'observer si l'hétérogénéité des interprétations possibles d'une même empreinte se retrouve pour les termes disparus.

B. Disparition de termes

Parallèlement à la liste de termes apparus dans le TTVS, la liste des 5 termes disparus (Tableau 4.8) a été présentée aux mêmes experts. Pour cette tâche, la question posée est la suivante : *les termes présentés sont-ils selon vous des termes obsolètes ou disparus du domaine de l'optique et de l'optoélectronique spatiales ? Sinon, pouvez-vous les lier à une évolution du domaine ?*

Rappelons que cette liste n'est pas significative d'un point de vue statistique dans la mesure où les termes sont trop peu fréquents. Elle permet néanmoins d'évaluer certaines tendances dans l'interprétation des experts. Dans cette liste, seuls 2 termes ont un χ^2 supérieur au seuil choisi pour cette recherche : *Modèle Instrumental* et *Modèle Radiométrique Instrumental*. Ce qui est intéressant dans cet exemple est qu'il s'agit de deux variantes du même terme qui disparaissent simultanément pouvant ainsi laisser penser que le concept est devenu obsolète. Or, les experts ont été unanimes : cette double disparition n'est pas due à l'obsolescence de ce terme/concept dans le domaine mais à la disparition de projets spatiaux cités en exemple dans le cours pour lesquels ce concept est central. En effet, ces deux termes sont utilisés dans le cadre des projets Spot 1 à 4 pour lesquels ont été mis en place des modèles instrumentaux spécifiques de simulation des performances radiométriques. Dans ces projets, un soin tout particulier a été porté à ces questions et les modèles instrumentaux développés pour Spot 1 à 4 constituent des références à citer dans le TTVS sur ces questions. Lors de la mise en place de la génération Spot 5, dès 1994, ces modèles sont devenus obsolètes et ont disparu du cours. Plus précisément, ce sont les modèles instrumentaux précis mis en place pour les projets Spot 1 à 4 qui sont potentiellement devenus obsolètes (bien que selon certains experts les modèles mis en place pour Spot servent encore de référence). Ce sont donc des instances de modèles instrumentaux qui ont disparu et non pas le terme ou le concept. Le concept *modèle*

instrumental n'est en effet quant à lui pas du tout obsolète et reste une question primordiale dans le domaine.

Le terme *LDEF* quant à lui est un exemple de disparition de projet : *LDEF* était un projet américain qui a embarqué le projet français FRECOPA en 1984. Si dans la première édition du TTVS, en 1994, on mentionne encore quelques retours d'expériences sur ce projet important, le temps passant ceux-ci deviennent obsolètes et doivent laisser place à de nouveaux projets de référence. *LDEF* constitue donc un exemple d'obsolescence d'un projet spatial.

La disparition du terme *synthèse d'ouverture optique* est un cas intéressant qui illustre un autre type de phénomène : la disparition de termes seuls. En effet, certains termes du domaine se trouvent remplacés par des variantes qui renvoient au même concept. C'est le cas du terme *synthèse d'ouverture optique* qui est remplacé petit à petit par le terme *synthèse d'ouverture*. Cet exemple illustre la dissociation que l'on peut être amené à faire en diachronie entre l'évolution du terme et l'évolution du concept. La proposition mentionnée au Chapitre 2 de dissocier le niveau du terme et le niveau du concept s'avère donc pertinente. Cet exemple est d'autant plus intéressant qu'il souligne la variabilité des interprétations possibles à partir d'un même indice : pour une même empreinte de fréquence, on passe de l'interprétation (souvent privilégiée) de la disparition du terme/concept à une interprétation d'évolution de la dénomination d'un concept. Il laisse également entrevoir l'existence d'autres indices linguistiques pour repérer l'évolution, ici la succession de variantes.

Enfin, l'exemple *instrument imageur* a été beaucoup plus difficile à traiter par les experts. Il s'agit en effet d'un terme très générique dans le domaine et il y a trop peu d'occurrences pour mener des analyses fines de contextes qui pourraient guider leur intuition. Il semble néanmoins là encore s'agir d'un changement de dénomination : on utilise de moins en moins le terme *instrument imageur* dans le cours, remplacé alors par des termes plus spécifiques tels que *radiomètres imageurs* ou par la variante réduite *imageurs*.

Bien que la variation de fréquence de ces 5 termes ne soit pas significative d'un point de vue statistique, les entretiens avec les experts tendent à confirmer le fait qu'une même empreinte

peut être associée à différentes interprétations. Ainsi, la disparition de ces termes peut être potentiellement liée à :

- un changement de dénomination (*synthèse d'ouverture, instrument imageur*),
- l'obsolescence du concept (*LDEF*),
- l'obsolescence d'une instance du concept (*modèle instrumental*).

Sur la base de ces premiers éléments, les mêmes entretiens sont mis en place sur le corpus DORIS. Comme indiqué au Chapitre 3 (§3.2.2.2, p.88), ce corpus est un corpus diachronique comparable, mais assez différent du corpus TTVS : il s'agit en effet d'un corpus projet, d'un degré de spécialité plus haut et découpé en deux sous-corpus seulement. Bien qu'il s'agisse d'un petit corpus, les observations de l'indice de fréquence permettent de compléter ou de nuancer les remarques établies dans le corpus TTVS.

4.3.1.2 Corpus DORIS

Le corpus DORIS est divisé en deux sous-corpus qui correspondent aux première et troisième générations de l'instrument DORIS. Dans le cas de l'hypothèse d'apparition, le corpus d'exclusion est donc le corpus le plus ancien (Gen1) et le corpus de référence est plus récent (Gen3) (Tableau 4.9). Dans le cas de l'hypothèse de disparition, le corpus d'exclusion est le plus récent et le corpus de référence est le plus ancien (Tableau 4.10).

Forme	Fréquence dans les sous-corpus	
	Gen1	Gen3
<i>terminal de transfert</i>	0	92
<i>TCH</i>	0	49
<i>télégestion</i>	0	39
<i>visibilité</i>	0	34
<i>IHM</i>	0	18

Tableau 4.9 – Exemples de termes apparus (DORIS)

Forme	Fréquence dans les sous-corpus	
	Gen1	Gen3
<i>touche</i>	68	0
<i>afficheur</i>	64	0
<i>appuyer</i>	49	0
<i>appuyer sur touche</i>	33	0
<i>afficheur indiquer</i>	25	0

Tableau 4.10 – Exemples de termes disparus (DORIS)

Au total dans le corpus DORIS, avant l'application du χ^2 , 301 termes apparaissent et 210 disparaissent. Par rapport au corpus TTVS, une nette différence se dessine : alors que 12% des termes extraits dans le TTVS sont concernés par les empreintes d'apparition/disparition, plus de 54% des 937 termes retenus dans le corpus DORIS apparaissent/disparaissent. De plus, alors que dans le TTVS une plus grande proportion de termes apparaît, les proportions sont beaucoup plus équilibrées dans DORIS : 59% des 511 termes apparaissent, contre 41% qui disparaissent. Dans la mesure où les deux sous-corpus DORIS sont équilibrés d'un point de vue quantitatif, on peut dire que la terminologie du projet DORIS, telle qu'elle est reflétée dans les rapports de spécification, reste assez équilibrée dans le temps. D'un point de vue qualitatif cependant, le fait qu'il y ait presque autant de termes qui disparaissent que de termes nouveaux laisse potentiellement envisager un certain « renouvellement » de la terminologie du projet.

Une fois le χ^2 appliqué⁷⁶, 18 termes apparus et 9 termes disparus sont jugés statistiquement significatifs. Comme les Tableau 4.11 et Tableau 4.12 l'illustrent, les valeurs du χ^2 obtenues sont beaucoup plus significatives que dans le cas du TTVS (Tableau 4.7). Ceci semble indiquer que les empreintes d'apparition/disparition dans le corpus DORIS sont beaucoup plus marquées que dans le cas du TTVS.

⁷⁶ Comme dans le cas du TTVS, et sur une suggestion de N. Ménard, les termes significatifs retenus le sont sur la base du χ^2 et d'une limite de fréquence pour assurer la significativité du test. Pour le corpus DORIS, la limite imposée est de 20 occurrences dans l'ensemble du corpus. Notons que sans cette limite de fréquence, 78 termes apparus et 67 termes disparus seraient significatifs selon le test du χ^2 « brut » dans DORIS.

Forme	DORISGen1	DORISGen3	χ^2
<i>terminal de transfert</i>	0	92	84,277
<i>blocs</i>	0	76	69,137
<i>configuration</i>	0	71	64,419
<i>CCD</i>	0	50	44,675
<i>TCH</i>	0	49	43,738
<i>contexte</i>	0	41	36,249
<i>télégestion</i>	0	39	34,38
<i>acquittements</i>	0	38	33,445
<i>TLG</i>	0	38	33,445
<i>fichier</i>	0	35	30,644
<i>visibilité</i>	0	34	29,711
<i>reçus</i>	0	33	28,778
<i>TLGP</i>	0	32	27,846
<i>module de gestion</i>	0	26	22,257
<i>logiciel</i>	0	22	18,539
<i>LSB</i>	0	22	18,539
<i>CCI</i>	0	22	18,539
<i>créneau</i>	0	21	17,611

Tableau 4.11 – 18 « Candidats néologismes » significatifs ($\chi^2 \geq 6,635$ et freq ≥ 20) – DORIS

Forme	DORISGen1	DORISGen3	χ^2
<i>touche</i>	68	0	72,043
<i>afficheur</i>	64	0	67,647
<i>appuyer</i>	49	0	51,206
<i>opérateur appuyer</i>	33	0	33,747
<i>appuyer sur touche</i>	33	0	33,747
<i>VAL</i>	32	0	26,136
<i>opérateur appuyer sur touche</i>	26	0	26,136
<i>afficheur indiquer</i>	25	0	25,05
<i>touche VAL</i>	22	0	21,795

Tableau 4.12 – 9 « Candidats obsolètes » significatifs ($\chi^2 \geq 6,635$ et freq ≥ 20) – DORIS

A. Apparition de termes

Sur la base de ces listes de termes, la même consigne que dans le cas du TTVS est proposée à l'expert DORIS impliqué dans cette tâche. La liste de 18 termes apparus présentée dans le Tableau 4.11 permet de dégager plusieurs tendances dont la diversité confirme l'hétérogénéité des interprétations observées pour le corpus TTVS. Rappelons également que quatre experts ont pu être interrogés sur ce type de tâche dans le TTVS alors qu'un seul est impliqué dans DORIS, limitant ainsi les possibilités d'un consensus et les retours et

discussions d'un expert à l'autre pour développer leur sentiment sur l'évolution (cf. Chapitres 9 et 10).

Dans cette liste, les termes *visibilité* et *créneau*, renvoient à l'apparition d'une nouvelle fonctionnalité présente dans l'instrument DORIS de troisième génération. En effet, entre 1990 et 2000, de nouveaux besoins sont apparus dans le spatial qui ont entraîné le développement d'une fonction d'émission dite sur *visibilité satellite*. Il s'agit d'un mode d'émission dans lequel la balise n'émet que lorsqu'elle rentre en contact avec certains satellites spécifiques (et non plus en continu comme cela était le cas dans la première génération). La balise n'émet alors que dans certains *créneaux* d'émission autorisée. Cet aspect de l'évolution de DORIS est important dans la mesure où il permet de souligner la « relativité » de la notion de nouveauté. En effet, si d'un point de vue terminologique la nouveauté est généralement relative à un domaine de connaissance et permet de travailler sur la néologie, dans ce contexte précis elle est relative à un projet. Cet aspect doit être mis en lien avec la notion de *domaine d'activité* proposée par De Bessé (2000) et présentée au Chapitre 2 (§2.1.1.3, p.45). En effet, le corpus DORIS est un corpus relatif à un projet, une activité précise. De fait, la nouveauté que l'on rencontre dans ce cadre peut n'être relative qu'à cette activité, comme dans l'exemple de *visibilité satellite*. Dans ce cas, il est difficile de parler de néologie dans la mesure où, du point de vue des connaissances générales dans le domaine spatial, ce terme/concept n'est pas néologique et peut se retrouver dans d'autres contextes plus anciens. Pour cette raison, dans la mesure où ces cas présents dans le corpus DORIS renvoient à la mise en place de fonctionnalités techniques, nous ne parlerons pas de néologie mais de l'apparition de nouvelles fonctionnalités.

Dans ce corpus, une autre tendance d'évolution s'est dégagée nettement : en effet, dans la liste des 18 termes proposés, 10 ont été mis en lien par l'expert avec les progrès de l'informatique. Ces 10 termes, *blocs*, *configuration*, *TCH*, *contexte*, *télégestion*, *TLG*, *fichier*, *reçus*, *TLGP*, *module de gestion* et *logiciel*, appartiennent à la terminologie de l'informatique et indiquent des progrès techniques et technologiques qui se répercutent dans le fonctionnement de l'instrument DORIS. Dans cette liste cependant, les termes/concepts *blocs*, *contexte* et *reçus* sont des termes très généraux et n'ont pu être traités par l'expert que dans la

mesure où les contextes dans lesquels ils apparaissent les lient avec les avancées informatiques. Par exemple, *blocs* est utilisé dans des syntagmes tels que *bloc de téléchargement*, *blocs de TCH*, *blocs à télécharger*, *blocs de télégestion*, *contextes* dans des syntagmes tels que *contextes de téléchargement*, *fichiers de contextes* et de même pour *reçus* (*téléchargements reçus*, *fichiers reçus*, *contextes reçus*, etc.). La présence de ces termes souligne la très forte expansion que connaît le domaine informatique, domaine qui revêt un caractère « transversal » : il s'agit d'un domaine technique mis à contribution dans de nombreux autres domaines. Les progrès de l'informatique se répercutent donc inévitablement dans d'autres domaines et applications. Le projet DORIS a lui aussi bénéficié des développements et progrès informatiques dont les bénéfices se retrouvent dans sa terminologie.

L'apparition du terme *terminal de transfert* fait émerger encore un autre type d'évolution : une évolution de la documentation du projet. À l'époque de la première génération DORIS, les documents de spécification ne détaillaient pas la question du *terminal de transfert* dans le corps des documents mais en annexes. Dans la mesure où il s'agit d'un élément central de l'instrument, dans la troisième génération cette question est ramenée dans le corps des rapports de spécification. De fait, ce terme renvoie à un concept qui existe depuis le début de la mise en place de DORIS, mais qui n'apparaît que dans les rapports de spécification de la 3^{ème} génération. L'évolution en jeu ici est donc due à la fois au choix du corpus construit pour cette recherche (en particulier le choix de ne pas sélectionner les annexes), ainsi qu'à un changement dans la structure des documents de référence du projet.

Enfin, l'apparition du terme *CCD* illustre un aspect également rencontré dans le TTVS : l'apparition de variantes de termes. Ainsi, l'apparition de *CCD* est en fait une variante du terme *CDC* préféré dans la génération 1 pour désigner le *centre de contrôle DORIS*.

Notons pour terminer que, pour 3 des termes apparus de la liste, l'expert n'a pu porter de jugement : *CCI*, *LSB* et *acquittements*. Pour chacun d'entre eux, il a semblé à l'expert qu'ils auraient pu apparaître dès la première génération, mais il n'a pu développer plus ce sentiment. Ceci est en partie dû au fait que pour l'ensemble des 18 termes présentés, quatre phénomènes d'évolution saillants ont pu être dégagés dans lesquels tous les termes ont pu

être regroupés et liés. Les trois termes non classés n'ont pu y être clairement associés, ce qui a mené l'expert à être prudent dans son jugement.

Bien que le corpus DORIS soit très différent du corpus TTVS, une même tendance se confirme : pour une même empreinte de fréquence, plusieurs interprétations d'évolution existent. Cependant, les interprétations proposées dans DORIS sont légèrement différentes de celles du TTVS. Dans le projet DORIS, l'apparition de termes peut être liée notamment à :

- l'apparition d'une variante (*CCD*),
- l'influence des progrès d'un domaine connexe (*téléchargement, fichiers, etc.*),
- la modification de la structure de documents de référence (*terminal de transfert*),
- l'apparition d'une nouvelle fonctionnalité de l'instrument (*visibilité, créneau*).

De plus, une autre caractéristique importante du corpus DORIS semble être la tendance des termes apparus à pouvoir être regroupés dans un même mouvement d'évolution. Mais avant de détailler ce phénomène, il convient de décrire les termes disparus afin d'observer si ces tendances se confirment dans le corpus.

B. Disparition de termes

Parallèlement à la liste des termes apparus, la liste des termes disparus est proposée à l'expert. Pour cette tâche, la même consigne a été donnée que pour le TTVS. Dans cette liste, l'expert a proposé de regrouper les termes en deux séries.

Une première série de termes réunit des éléments obsolètes disparus sur l'instrument. Dans ce cas, les concepts renvoient à des éléments de la génération 1 que l'on ne retrouve pas dans la génération 3 : *touche, afficheur, touche VAL*. Comme dans le cas de *Modèle Instrumental* décrit *supra* pour le TTVS, le concept *touche* n'est pas obsolète, mais l'instance présente sur la première génération l'est : il n'y a plus de *touche VAL* sur les nouvelles balises. Dans le cas d'*afficheur* par contre, c'est bien le concept qui est obsolète, remplacé par le concept plus moderne *écran* (cf. §7.3.4, p.243 et §9.1.3, p.281).

La seconde série proposée regroupe des termes disparus qui peuvent être mis en lien avec la série précédente. En effet, la disparition d'éléments obsolètes entraîne la disparition d'actions

ou d'événements qui leur sont liés. Ces actions ne peuvent plus être réalisées et disparaissent à leur tour de la terminologie du projet DORIS. Il s'agit de *appuyer*, *opérateur appuyer*, *appuyer sur touche*, *opérateur appuyer sur touche*, *afficheur indiquer*.

Ces deux regroupements, dégagés nettement par l'expert interrogé, indiquent que dans le corpus DORIS les termes *apparus/disparus* tendent bien à évoluer de concert, élément que l'on n'a pu observer dans le corpus TTVS, pourtant plus gros. Encore une fois, les interprétations d'évolution sont multiples et nous pouvons mettre en lien la disparition de termes du corpus DORIS avec notamment :

- la disparition d'instances obsolètes (*touches*),
- la disparition de concepts obsolètes (*afficheur*),
- la disparition d'actions ou d'événements liés à la disparition d'instances ou de concepts (*appuyer*, *position*).

4.3.1.3 Synthèse

Les descriptions proposées pour chacun des corpus permettent de souligner plusieurs aspects importants quant à l'analyse de l'évolution des connaissances à partir de corpus.

Le premier élément à souligner, et qui n'est que rarement mentionné dans la littérature, est la multiplicité des interprétations d'évolution à partir de corpus. Dans le cas de l'indice de fréquence, pour une même empreinte, les experts ont pu associer différentes interprétations telles que la néologie, le rappel de termes/concepts anciens pour décrire des évolutions, l'apparition de nouvelles fonctionnalités, le développement d'un domaine connexe, l'apparition de variantes de termes, la mise en place de projets, la modification de la structure des documents, etc. De plus, dans la mesure où les observations menées dans cette recherche ne le sont que sur deux corpus assez petits et – à ce stade de notre étude – sur un seul indice, ces diverses interprétations pourront être complétées par d'autres aspects susceptibles d'intervenir dans l'évolution des connaissances d'un domaine.

Le deuxième élément est qu'il s'avère nécessaire de dissocier évolution de la dénomination et évolution du concept. En effet, plusieurs exemples ont montré que la dénomination peut

évoluer sans que cela implique un changement conceptuel. De la même manière, à l'évolution du terme et du concept doit être ajouté un troisième niveau d'évolution : celui des instances de concepts. En effet, dans les cas de *Modèle Instrumental* (TTVS) et *touche VAL* (DORIS) il ne s'agit ni de la disparition/obsolescence d'un terme/concept ni de celle d'une dénomination mais de la disparition/obsolescence de l'instance d'un concept présente dans un projet ou sur un instrument donné.

Le troisième élément que l'on peut établir est que le type de corpus semble avoir une influence sur les observations possibles. En effet, la comparaison de l'indice de fréquence entre le corpus TTVS et le corpus DORIS montre au moins deux différences. Premièrement, l'indice de fréquence semble beaucoup plus marqué dans le corpus technique DORIS que dans le corpus didactique TTVS : les valeurs χ^2 sont plus fortes et les termes/concepts apparus et disparus sont susceptibles d'être regroupés massivement par les experts dans des séries d'évolution homogènes. Deuxièmement, les évolutions repérables dans DORIS sont principalement des évolutions techniques et fonctionnelles dans la mesure où la représentativité du corpus ne se situe pas au niveau d'un domaine de connaissance (comme c'est le cas pour le TTVS) mais au niveau d'un domaine d'activité. Nous l'avons vu avec l'exemple de *visibilité satellite*, cet aspect est susceptible de se répercuter sur la question de la néologie : il pourra en effet être plus difficile de parler de néologie dans un corpus projet où les évolutions se concentrent sur un domaine d'activité que dans un corpus représentatif d'un domaine de connaissance.

Quoiqu'il en soit, à ce niveau, on ne dispose que de trop peu d'informations pour véritablement généraliser ces éléments, mais ils constituent des pistes et des tendances très riches à partir desquelles poursuivre l'observation.

Pour finir, revenons sur un autre élément essentiel dans notre recherche : la collaboration avec des experts dont nous avons questionné ici le sentiment sur des termes identifiés en corpus grâce aux empreintes de fréquence. À travers ces premiers entretiens, une tendance est apparue : dans le cas du TTVS, où il n'est pas possible de regrouper massivement les termes en séries d'évolution, les experts ont été obligés de proposer des analyses locales pour chacun des termes isolés de la liste. Cette manière de proposer une interprétation est

beaucoup plus délicate pour eux que dans le cas de DORIS où les termes sont regroupables en séries. En effet, le fait de pouvoir traiter des séries de termes permet de proposer une interprétation globale sur un mouvement d'évolution et d'asseoir ainsi leur sentiment sur l'évolution du domaine.

Pour tenter de développer cet aspect et offrir un second point de vue sur l'indice de fréquence, la prochaine section propose l'analyse de formes chrono-homogènes dont l'intérêt est précisément de permettre de regrouper des termes dont les empreintes de fréquence sont comparables pour en rechercher ensuite les causes externes possibles d'évolution (Habert, *et al.*, 1997). Ce point de vue offre un deuxième angle d'approche méthodologique pour aborder le traitement de l'évolution des connaissances en corpus.

4.3.2 Termes chrono-homogènes

4.3.2.1 Repérage de thématiques

Le point de vue sur l'indice de fréquence adopté dans cette section repose à la fois sur le repérage d'empreintes de fréquence continues (vs. rupture apparition/disparition) et sur la recherche de groupes de plusieurs termes qui s'inscrivent ensemble dans une même empreinte de fréquence dans le temps. Ces termes sont appelés « chrono-homogènes ». À partir des groupes de termes ainsi formés, l'objectif est de chercher à saisir si ces termes peuvent être regroupés en thématiques dont l'évolution dans le domaine peut être retracée. Dans ce type de tâche, les experts sont donc invités à la fois à regrouper les termes en thématiques et à partager leur sentiment d'évolution sur ces thématiques dans le domaine.

Les termes chrono-homogènes sont extraits ici sur le modèle décrit au paragraphe 4.1.3 (p.109). Les termes retenus pour former les groupes sont sélectionnés à partir du seuil de χ^2 décrit qui assure la significativité de leurs variations d'un point de vue statistique. Ces termes sont ensuite regroupés sur la base des écarts observés (§4.2.2, p.111). Ils sont répartis ensuite en 6 empreintes pour le TTVS (Tableau 4.13) :

Empreintes	Termes concernés	Nombre termes
1. Croissance à partir de 1998	<i>absorption, alignement, astigmatisme, cavité, cavité étendue, courbure, HRG, miroir, raie</i>	9
2. Croissance à partir de 2002	<i>APS, arêtes, baffle, cœur, colle, couches, déformation, diffusion, domaine spectral, face, fibre, image, milieu, mode, MOS, optique, parasites, photo, pixel, plan, polissage, TDI, transfert des charges, transition, utilisation</i>	26
3. Décroissance à partir de 1998	<i>antenne, atmosphère, bruit, cible, dimensionnement radiométrique, espace froid, fondeur, impulsion, instrumental, luminance, polder, reçu, refroidir, refroidissement, signal, solaire, vol</i>	17
4. Décroissance à partir de 2002	<i>fixation, GEO, LEO</i>	3
5. Pic de croissance en 1998	<i>aberrations, microns, concave</i>	3
6. Pic de décroissance en 1998	<i>système optique</i>	1

Tableau 4.13 – 6 empreintes de fréquence (TTVS)

Dans la mesure où le corpus DORIS ne compte que 2 sous-corpus, les termes y sont regroupés en 2 empreintes (Tableau 4.14) :

Empreintes	Termes concernés	Nombre termes
1. Croissance	<i>terminal, satellite, transfert, terminal de transfert, bloc, configuration, ccd, tch, contexte, télégestion, gestion, tlg, acquittement, réception, fichier, visibilité, recevoir, tlgp, donnée, module de gestion, téléchargement, émission, logiciel, lsb, cci, date, créneau, mot, max, synthétiseur, anomalie, site, passage, survie, interface, délai, identification, module, exigence, test, bord, période, autoriser</i>	49
2. Décroissance	<i>afficheur, afficheur indiquer, appuyer, appuyer sur touche, autotest, batterie, carte, clavier, clé, commande, connecteur, externe, face, fonctionner, fréquence, gamme, gamme de mesure, horloge, humidité, indique, indiquer, jour, mesure, mode programmation, opérateur, opérateur appuyer, opérateur appuyer sur touche, opération, panne, position, pression, programmer, puissance, recalage, résolution, second, seconde, secteur, séquençement, signal, sortir, stabilité, température, tension, touche, touche val, val, valeur, veille</i>	46

Tableau 4.14 – 2 empreintes de fréquence (DORIS)

Ces listes de termes sont ensuite présentées à deux des quatre experts avec la consigne suivante : *Dans chacune des listes qui vous sont proposées, vous est-il possible de regrouper les termes en fonction de thématiques communes ? Si oui, est-il possible de les mettre en lien avec une évolution du domaine ?*

Dans les paragraphes suivants les regroupements et interprétations proposés par les experts sont analysés pour chacun des corpus.

4.3.2.2 Interprétations des experts

A. Thématiques identifiées dans le corpus TTVS

Les thématiques dégagées unanimement par les experts pour le corpus TTVS sont présentées dans le Tableau 4.15 :

- la rubrique « Termes classés » contient les termes pris en compte par les experts pour la mise en place des thématiques,
- « Termes supprimés » renvoie aux termes que les trois experts interrogés ont jugé trop généraux ou non pertinents par rapport au domaine de l'optique spatiale,
- « Termes non classés » renvoie aux termes du domaine que les experts n'ont pas pris en compte pour l'établissement des thématiques.

Comme l'illustre le tableau suivant, une à deux thématiques ont pu être identifiées par les experts pour chacun des groupes de termes chrono-homogènes⁷⁷. Ces thématiques englobent une proportion importante des termes dans chaque groupe puisque de 70 à 90% des termes présentés dans les groupes sont classés par les experts. Seuls les termes en décroissance à partir de 1998 (groupe 5) ne sont classés qu'à hauteur de 53%. Ceci permet de confirmer l'intérêt d'une démarche globale pour l'observation de l'évolution : une proportion considérable de termes du domaine montre une tendance certaine à évoluer de concert dans les corpus.

⁷⁷ L'empreinte « Pic de croissance en 1998 » ne compte qu'un seul terme, *système optique*. Dans la mesure où il ne s'agit que d'un terme isolé pour lequel les experts n'ont pu proposer d'interprétation, il n'est pas repris dans la suite de cette section (et notamment dans le Tableau 4.15.).

	Empreintes	Nombre de termes	Termes classés	Termes non classés	Termes supprimés	Nombre de thématiques	Thématiques	Termes classés	Termes non classés	Termes supprimés
1	Croissance à partir de 2002	26	23 (88%)	3	0	2	Technologies de fabrication et intégration mécano-optique des miroirs	colle, cœur, baffle, fibre, polissage, milieu, arêtes, face, plan, diffusion, parasites, couches, déformation		
							Techniques de détection /d'observation	APS, optique, domaine spectral, pixel, TDI, photo, MOS, image, transfert des charges	mode, utilisation, transition	
2	Croissance à partir de 1998 jusqu'en 2002	9	8 (89%)	0	1	2	Instruments lasers	cavité, raie, cavité étendue, absorption		
							Techniques d'imagerie et systèmes optiques	alignement, astigmatisme, courbure, miroir	HRG	
3	Décroissance à partir de 2002	3	2 (67%)	1	0	1	Orbites et satellites	GEO, LEO		fixation
4	Décroissance à partir de 1998 jusqu'en 2002	17	9 (53%)	6	2	1	Détection à basse température (IR)	espace froid, refroidir, luminance, bruit, signal, antenne, dimensionnement radiométrique, instrumental, refroidissement	fondeur, reçu, cible, solaire, impulsion, polder	atmosphère, vol
5	Pic de croissance en 1998	3	2 (67%)	1	0	1	Techniques optiques d'imagerie (des miroirs)	aberrations, axe optique	micron	

Tableau 4.15 – Répartition des termes chrono-homogènes en thématiques (TTVS)

Dans l'ensemble, les experts ont pu dégager assez aisément des thématiques dans ces groupes de termes, thématiques qui se révèlent assez précises pour le TTVS. De plus, elles reflètent des évolutions pertinentes du domaine de l'optique et optoélectronique spatiales. Seule la thématique « Orbite et satellites » regroupant les termes en décroissance à partir de 2002 (Groupe 5) n'a pu être liée à une évolution du domaine. La raison tient au fait que ce groupe contient très peu de termes (3 au total).

Pour chaque groupe, la méthodologie adoptée par les experts est une description descendante, du niveau global des thématiques au niveau local des termes isolés. Les entretiens se sont déroulés en trois phases :

1. présentation des listes de termes aux experts et identification de thématiques,
2. jugement de la pertinence des thématiques du point de vue de l'évolution des connaissances et discussion sur ces évolutions,
3. amorce d'observations locales de termes présents dans les thématiques qui pourront ensuite être complétées.

Afin d'illustrer la démarche descendante que permet l'analyse des termes chronologiques, les éléments principaux que les entretiens ont fait émerger pour chacune des thématiques sont rapportés dans la suite de cette section.

Pour le premier groupe, « Croissance à partir de 2002 », deux thématiques ont été identifiées. La première, intitulée « Technologies de fabrication et d'intégration mécano-optique »⁷⁸, renvoie à une série de termes liés à la fabrication et à l'intégration des miroirs. Cette thématique renvoie au développement des questions liées à la fabrication des miroirs en optique. Le terme/concept *miroir* n'est ni nouveau ni récent dans le domaine (puisque les miroirs participent aux fondements mêmes des techniques optiques) et n'apparaît d'ailleurs pas dans cette thématique. Cependant, les questions telles que celles des colles, du polissage et des parasites, qui apparaissent dans la série, sont liées à des développements importants dans le domaine. Nous reviendrons sur ces questions dans la suite de notre recherche.

⁷⁸ Les thématiques sont nommées par les experts qui les ont dégagées.

La seconde thématique est intitulée « Technique de détection/d'observation ». Les experts ont eu beaucoup plus de difficulté à l'associer à des évolutions du domaine. En effet, dans le spatial, l'optique est quasiment toujours impliquée dans des tâches de détection et d'observation, questions sur lesquelles le TTVS met particulièrement l'accent. Il n'est donc pas surprenant de retrouver ce type de thématiques en regroupant des termes du TTVS. De plus, les termes rassemblés dans ce groupe sont assez généraux (par exemple *optique*, *image*, *photo*) et il a été difficile aux experts de les rattacher à des évolutions précises. De manière globale, ces derniers s'accordent sur le fait que les techniques de détection et d'observation évoluent beaucoup, mais peu de précisions peuvent être apportées à ce sentiment. D'un point de vue local cependant, et parallèlement à ce qui a été dit *supra* lors de l'observation de termes apparus, le terme *APS* peut attirer l'attention puisqu'il s'agit d'un des rares néologismes relevés dans le corpus.

Pour le deuxième groupe, « Croissance à partir de 1998 », deux thématiques ont également été identifiées. La première, « Instruments lasers », contient peu de termes/concepts mais renvoie à une thématique que les experts désignent comme étant sujette à des évolutions. Cette catégorie doit être mise en lien avec l'analyse proposée par les experts dans le cas du terme/concept *cavité étendue* (cf. *supra*) et confirme l'interprétation selon laquelle les technologies associées aux lasers évoluent dans le spatial depuis le milieu des années 1990.

La seconde thématique, intitulée « Techniques d'imagerie et systèmes optiques », est mise en lien avec la thématique « Technologies de fabrication et d'intégration mécano-optique » identifiée pour le groupe de termes en croissance à partir de 2002 et concerne les progrès liés aux miroirs et à différentes techniques qui les touchent (dont les colles, la question des parasites, du polissage, etc.). Le fait que ces deux thématiques puissent être associées permet de renforcer la pertinence de l'interprétation selon laquelle plusieurs questions liées aux miroirs sont développées dès 1998. Cet aspect montre à la fois la pertinence de cette démarche à partir de thématiques, mais également la cohérence des analyses des experts.

Pour les trois groupes restants, une seule thématique a pu être mise au jour pour chacun.

La thématique « Orbite et satellite » du troisième groupe (Décroissance à partir de 2002) ne peut être associée à une évolution du domaine : les experts interrogés n'ont pu trouver d'arguments pour lier cette question à une évolution du domaine.

De la même manière, le groupe 5 « Pic de croissance en 1998 » ne contient que très peu de termes et il est difficile de les répartir en thématiques. Cependant, au vu des deux thématiques identifiées précédemment, qui mettent en avant l'évolution des questions associées aux miroirs, les experts ont pu regrouper deux de ces termes sur une thématique similaire « Techniques optiques d'imagerie (des miroirs) ».

Enfin, la thématique « Détection à basse température » (groupe 4, Décroissance à partir de 1998) est particulièrement intéressante dans la mesure où elle se révèle très précise. Il s'agit en effet de l'une des thématiques pour laquelle les experts ont montré le moins d'hésitation et où le consensus est le plus net. Cet aspect explique certainement le fait que la proportion de termes qui y sont classés soit la plus faible parmi tous les groupes (53% au lieu des 70 à 90% pour les autres). En effet, plus la thématique est précise et claire pour les experts, plus il est facile pour eux de sélectionner les termes/concepts pertinents, sans hésiter. À l'inverse, lorsque les thématiques sont générales ou plus discutables les experts ont tendance à rajouter plus de termes, au risque de mettre au jour des thématiques « fourre-tout » rendant difficile l'observation de phénomènes pertinents (c'est le cas par exemple de la thématique « Techniques de détection /d'observation » du groupe 1). D'un point de vue évolutif, cette thématique particulièrement précise reste à être développée : elle semble indiquer en effet que la thématique du refroidissement soit moins présente dans les éditions les plus récentes. Nous détaillons plus loin l'analyse de cette question, qui est l'une des évolutions les plus marquées et les plus riches dans le TTVS (cf. §9.1.2, p.279).

B. Thématiques dans le corpus DORIS

Pour compléter les observations possibles dans le cadre de l'analyse des termes chronohomogènes, la même démarche est proposée sur le corpus DORIS. Les thématiques dégagées pour ce corpus sont présentées dans le Tableau 4.16. Ce tableau est identique à celui proposé pour le corpus TTVS (p.135) ; seule la catégorie « Termes supprimés », restée vide, n'a pas été reprise.

Rappelons néanmoins que la tâche d'identification des thématiques n'a été effectuée que par un expert (*vs.* deux pour le corpus TTVS). De fait, celles-ci n'ont pu être rediscutées ; une plus grande proportion de termes reste donc non classée (de 36 à 50% de termes classés au maximum), faute de discussion et de temps essentiellement.

	Empreintes	Nombre de termes	Termes classés	Termes non classés	Nombre de thématiques	Thématiques	Termes classés	Termes non classés
1	Croissance	43	22 (51%)	21	2	Informatique	<i>bloc, configuration, tch, contexte, télégestion, gestion, tlg, réception, fichier, recevoir, tlgp, donnée, module de gestion, téléchargement, émission, logiciel, interface</i>	<i>terminal, lsb, cci, date, créneau, mot, max, synthétiseur, anomalie, site, survie, acquittement, ccd, transfert, terminal de transfert, délai, module, exigence, test, bord, période,</i>
						Fonction de visibilité satellite	<i>passage, visibilité, satellite, identification, autoriser</i>	
2	Décroissance	49	13 (36%)	36	1	Changements sur la forme de la balise	<i>afficheur, afficheur indiquer, appuyer, appuyer sur touche, clavier, clé, opérateur, opérateur appuyer, opérateur appuyer sur touche, position, touche, touche val, val</i>	<i>connecteur, externe, face, fonctionner, fréquence, gamme, gamme de mesure, horloge, humidité, indique, indiquer, jour, mesure, mode programmation, autotest commande, batterie, carte, opération, panne, pression, programmer, puissance, recalage, résolution, second, seconde, secteur, séquençement, signal, sortir, stabilité, température, tension, valeur veille</i>

Tableau 4.16 - Répartition des termes chrono-homogènes en thématiques (DORIS)

L'identification de thématiques dans le corpus DORIS présente des résultats très homogènes par rapports aux observations formulées dans la section 4.3.1.2 (p.124) : les thématiques observées sont très nettes et reprennent les observations proposées par l'expert sur les premières listes de termes *apparus/disparus*. Le fait que les thématiques fassent écho aux premières observations proposées *supra* s'explique de plusieurs manières. Le premier élément est que, les interprétations n'étant menées que par un seul et même expert sur le corpus DORIS, ses analyses restent cohérentes sur le même corpus. Le second élément est que les termes *apparus/disparus* du corpus DORIS ont montré une tendance forte à pouvoir être regroupés en séries pour être analysés. De fait, dans la mesure où l'identification de thématiques vise à exploiter cette tendance de regroupement, on retrouve dans le corpus DORIS une part des observations déjà posées. Cependant, la cohérence des interprétations laisse affirmer que l'évolution des connaissances dans le projet DORIS est également très homogène et se dessine à travers de grandes tendances que les thématiques font ressortir.

Les deux premières tendances sont définies à travers les deux thématiques proposées pour le premier groupe (termes *chrono-homogènes* en croissance) : une thématique « informatique » et une thématique « visibilité satellite », déjà repérées à travers l'analyse des termes *apparus/disparus* :

- la terminologie du projet DORIS est marquée par la migration massive de termes informatiques et les progrès techniques dans l'informatisation de la balise ;
- une nouvelle fonction est intégrée à la balise et marque une évolution profonde entre la première et la troisième génération de balise : l'émission sur visibilité satellite plutôt qu'une émission en continu. Ce dernier changement est d'autant plus important d'un point de vue externe qu'il est lié à l'émergence de nouveaux besoins du Cnes dont :
 - des besoins d'économie d'énergie : en effet, parmi l'ensemble des satellites qui embarquent l'instrument DORIS tous n'ont pas la même fonction. Certains ont des besoins pour la mesure de hauteurs altimétriques, d'autres sont dédiés à des « missions localisation ». De fait,

les balises qui ne participent qu'à la localisation n'ont pas besoin d'émettre sur des satellites qui ne participent pas à cette mission ;

- des besoins politiques : le Cnes est amené à mettre en place des accords de coopération internationale avec des pays dont certains imposent de ne pas utiliser la balise DORIS sur leur sol pour des missions qu'ils ne connaissent pas ou qu'ils suspectent d'être à but militaire. Le mode d'émission sur visibilité satellite est donc une réponse adéquate à ce type de besoins.

La troisième thématique identifiée concerne le dernier groupe (Décroissance) et est intitulée par l'expert « Changements sur la forme de la balise ». Cette catégorie contient un ensemble de termes/concepts qui renvoient à des concepts et des instances (cf. p.129) obsolètes sur la balise. Ceci reflète d'une part la modernisation du boîtier de la balise (suppression des touches, de la clé, etc.), mais également l'informatisation de certaines fonctions associées à ces éléments obsolètes (par exemple, on lance un programme au lieu de tourner une clé).

4.3.2.3 Synthèse

L'analyse de termes chrono-homogènes proposée dans cette section offre un second point de vue sur l'observation des fréquences de termes pour traiter de l'évolution des connaissances. Cette approche, menée sur les deux corpus d'étude, revêt plusieurs intérêts.

Tout d'abord, les observations menées dans cette section viennent confirmer les tendances relevées lors de l'observation des termes apparus/disparus (§01.3, p.115). L'analyse en thématiques montre en effet une fois encore l'hétérogénéité des aspects d'évolution susceptibles d'intervenir en diachronie courte et repérable en corpus. Les interprétations proposées lors de cette tâche sont notamment :

- pour le TTVS :
 - le développement de questions liées aux miroirs,
 - le développement des lasers,
 - la disparition de termes/concepts liés à la détection à basse température,

- pour DORIS :
 - le développement et la migration des techniques informatiques dans le projet,
 - la modernisation du boîtier de la balise,
 - l'apparition d'une nouvelle fonctionnalité.

La seconde forte tendance relevée est la différence de « comportement » entre les corpus DORIS et TTVS. Le corpus technique DORIS semble marquer plus nettement les différentes évolutions qui touchent le projet. En effet, les thématiques et évolutions identifiées dans le corpus le sont très nettement et viennent appuyer les observations décrites au §4.3.1.2.

À l'inverse, le corpus TTVS s'est révélé un peu plus difficile à traiter, bien que certaines thématiques soient très pertinentes et permettent d'amorcer une importante description de l'évolution (notamment sur la question des miroirs et du refroidissement).

Mais le plus grand intérêt de cette démarche est méthodologique. L'identification de thématiques à partir de groupes de termes chrono-homogènes permet de mettre en place une démarche descendante sur les données : d'une vision globale sur des groupes de termes, l'analyse s'affine peu à peu en une description locale des termes qui composent les thématiques. Il s'agit donc d'un moyen d'entrée tout à fait pertinent pour aborder l'étude d'un domaine dont on ne connaît pas l'évolution *a priori* : cette démarche permet de rechercher des classes saillantes de termes qui globalement évoluent ensemble, afin de repérer des pistes d'évolution à analyser plus finement. L'intérêt de cette approche réside donc dans l'articulation des niveaux de description globaux et locaux, articulation opérée à la fois par le linguiste et par l'expert et qui permet de ne poser qu'un minimum d'*a priori* sur les évolutions de connaissances du domaine. De plus, il est beaucoup plus aisé pour eux de poser un jugement sur l'évolution de groupes de termes/concepts que sur termes/concepts isolés. Ces groupes favorisent la reconstruction de liens entre les termes du groupe pour retracer leur évolution et rassurer ainsi l'expert dans son interprétation.

Il est évident que la place accordée à l'interprétation des données reste primordiale dans cette approche, puisque ce sont les experts qui proposent des thématiques sur la base de groupes de termes repérés par le linguiste. Dans cette démarche, l'identification des thématiques reste subjective et il est possible que les thématiques repérées divergent d'un expert à l'autre. De la même manière, certaines thématiques peuvent ne pas être pertinentes, se révéler trop générales (voire « fourre-tout »). Dans ce cas, elles ont été identifiées par les experts afin de ne pas rester sans réponse devant la tâche proposée et non pas parce qu'elles sont réellement pertinentes. Néanmoins, ce cas de figure ne s'est quasiment pas présenté dans nos recherches et la tendance des termes à pouvoir être regroupés en groupes chrono-homogènes s'est plutôt révélée un atout pour guider la tâche d'interprétation des experts. De plus, dans les cas où les experts de domaine ne sont pas ou très peu disponibles pour accompagner le linguiste dans sa tâche d'exploration de l'évolution, les regroupements de termes peuvent permettre à ce dernier d'entamer son analyse de manière autonome. En effet, puisque dans l'exemple des deux corpus d'étude une proportion considérable de termes chrono-homogènes peut être classée en thématiques, ils offrent un champ d'exploration riche et pertinent pour entamer une analyse en corpus avec un recours minimal aux experts (bien que le recours aux experts, même minimal, reste nécessaire).

4.4 Bilan

Dans ce chapitre, nous avons proposé plusieurs manières d'aborder la question de la fréquence en corpus comparables dans une perspective diachronique et méthodologique, en opposant à la fois les notions de rupture et de continuité, ainsi que les notions d'analyse locale et globale.

Les notions de rupture et de continuité se complètent dans la mesure où, en langue de spécialité, la néologie occupe généralement une place centrale des études sur l'évolution. Or, nos résultats montrent qu'en diachronie courte l'évolution est hétérogène et ne peut être « réduite » à l'existence de néologismes. Il est donc nécessaire d'associer différentes empreintes de fréquence pour une analyse exhaustive.

Les points de vue local et global revêtent un intérêt particulier grâce au fait qu'ils peuvent être articulés pour compléter l'analyse. En effet, la perspective globale permet de faire émerger des tendances marquées d'évolution, à l'aide de grandes classes de termes chronologiques homogènes. Une fois ces classes obtenues, une analyse locale est envisageable, qui permet d'observer l'évolution de chacun des termes ainsi regroupés et de tenter de retrouver les causes d'évolution qui expliquent l'homogénéité de leur variation. L'intérêt méthodologique de ce type de démarche est donc triple :

- dans le cas où le repérage cible une évolution connue *a priori* (par exemple l'apparition d'un nouveau concept) l'articulation entre analyses globales et locales permet d'affiner l'analyse de l'évolution et d'ouvrir la description de termes isolés (souvent privilégiée en diachronie courte) à la prise en compte de « réseaux » de termes qui évoluent ensemble ;
- dans le cas où aucune évolution n'est définie *a priori* (comme c'est le cas dans notre recherche), ce type de démarche permet d'amorcer l'observation de l'évolution pour n'importe quel domaine, même si l'on n'est pas expert, à condition de disposer de corpus comparables. On peut ensuite affiner la description des tendances observées à l'aide d'analyses locales et l'aide des experts ;
- dans tous les cas, d'un point de vue méthodologique, l'articulation entre points de vue local et global permet d'aider les experts et l'analyste à émettre un jugement sur l'évolution, dans la mesure où poser un jugement sur des séries de termes homogènes est généralement plus aisé que sur des termes isolés directement.

Dans ce chapitre, nous avons fait le choix de présenter de manière détaillée les différentes descriptions de termes présentés aux experts. Cette présentation, certainement un peu fastidieuse, est nécessaire afin de montrer la pertinence des empreintes de fréquence pour aborder la question de l'évolution, ainsi que la diversité des interprétations que l'on peut associer à un même indice. Cet élément est loin d'être trivial et mérite une description détaillée, d'autant plus que les empreintes de fréquence sont le premier indice traité dans

notre travail et, de fait, l'indice qui apporte les premiers éléments d'analyse pour construire la suite de notre recherche.

Les observations obtenues n'ont été possibles que grâce à la place privilégiée accordée aux experts et à l'interprétation « qualitative » des résultats obtenus. Elle permet de souligner deux phénomènes centraux. Le premier est que l'évolution d'un domaine peut effectivement se manifester sur des intervalles de temps très courts, bien que cela soit assez peu souvent admis. L'évolution mise au jour y est même très riche dans nos deux corpus.

Le second intérêt méthodologique des empreintes de fréquence est la possibilité de les mettre en œuvre dans n'importe quel type de corpus. Les phénomènes observés d'un corpus à l'autre peuvent cependant se révéler différents. Par exemple, les empreintes de fréquence sont apparues beaucoup plus marquées dans DORIS que dans le TTVS. Les évolutions mises au jour sont également plus homogènes dans DORIS que dans le TTVS, corpus qui révèle de nombreuses facettes d'évolution.

Néanmoins, une tendance se dégage qui permet d'envisager des pistes intéressantes pour observer l'évolution en corpus : la multiplicité des interprétations possibles s'accompagne d'une multiplicité d'indices linguistiques en corpus. Dans ce chapitre, cet aspect a été illustré par exemple à travers l'apparition/disparition conjointe de variantes (par ex. *CCD/CDC* dans DORIS), ou encore dans DORIS par le fait que le développement de progrès informatiques peut causer, du moins en partie, la disparition d'éléments sur la balise. Ce phénomène n'est qu'effleuré dans ce chapitre, mais ouvre une perspective intéressante quant à l'identification et à la description de nouveaux indices d'évolution en corpus, ce que nous poursuivons dans le chapitre suivant.

Chapitre 5 Contextes riches en connaissances évolutives

SOMMAIRE DU CHAPITRE

5.1	CONTEXTES RICHES EN CONNAISSANCES ÉVOLUTIVES ET MARQUEURS D'ÉVOLUTION	148
5.1.1	<i>Définition.....</i>	148
5.1.2	<i>Exemples d'extraction d'informations évolutives dans les textes</i>	149
5.2	EXTRACTION DES CONTEXTES D'ÉVOLUTION À L'AIDE DE MARQUEURS.....	158
5.2.1	<i>Marqueurs définis pour cette recherche.....</i>	158
5.2.2	<i>Extraction de contextes riches en connaissances évolutives</i>	168
5.3	OBSERVATIONS EN CORPUS : PROJECTION DES TERMES EN CONTEXTES	169
5.3.1	<i>Évaluation de la pertinence des contextes et des marqueurs</i>	169
5.3.2	<i>Observations de contextes</i>	177
5.3.2.1	Point de vue synchronique dynamique.....	178
5.3.2.2	Point de vue diachronique	180
5.3.3	<i>Observations de termes récurrents en contextes.....</i>	182
5.4	BILAN.....	185

Nous présentons dans ce chapitre un deuxième indice à travers la définition et le repérage de contextes riches en connaissances évolutives. Ce type d'indice s'inspire de techniques et d'approches d'extraction d'information et de fouille de textes, basées sur des analyses locales d'éléments textuels (appelés ici marqueurs) pour repérer des éléments d'informations pertinents en corpus.

Dans un premier temps, nous définissons la notion de « contextes riches en connaissances évolutives » pour notre perspective. Dans un second temps, nous décrivons la méthodologie semi-automatique de repérage de ces contextes, sur la base de l'identification de marqueurs lexico-syntaxiques d'évolution. Nous terminons par une description détaillée des évolutions repérées et des analyses que l'on peut construire à l'aide de cet indice.

5.1 Contextes riches en connaissances évolutives et marqueurs d'évolution

5.1.1 Définition

Les *contextes riches en connaissances évolutives* sont entendus ici comme des portions de textes qui contiennent des informations pertinentes sur l'évolution des connaissances d'un domaine. Cette notion s'inspire de la notion de *contexte riche en connaissances* (ou *Knowledge-Rich Context*, KRC) proposée notamment par Meyer (2001) et que l'auteur définit comme

« indicating at least one item of domain knowledge that could be useful for conceptual analysis. In other words, the context should indicate at least one conceptual characteristic, whether it be an attribute or a relation. » (*ibidem* : 281)

Cette notion a vu le jour en terminologie et en ingénierie des connaissances et fait appel à la notion de *marqueurs* pour la caractérisation des contextes. Les marqueurs sont au cœur des tâches de construction et de structuration de ressources termino-ontologiques à partir de textes où l'« on suppose que l'observation de certaines formes linguistiques entre deux ou plusieurs éléments du lexique peut révéler un rapport de sens entre ces éléments » (Aussenac-Gilles & Séguéla, 2000 : 175). Ces formes linguistiques, lorsqu'elles permettent de rendre compte de fonctionnements stables et associables systématiquement à une interprétation sémantique, sont appelées *marqueurs de relations*⁷⁹. Les contextes dans lesquels apparaissent des marqueurs et des termes qui entrent dans la relation décrite sont alors appelés *contextes riches en connaissances*.

Dans la lignée de ces travaux, nous faisons l'hypothèse qu'il est possible de définir des marqueurs d'évolution pour identifier des contextes riches en connaissances évolutives, contextes qui contiennent des informations pertinentes sur l'évolution du domaine.

⁷⁹ Voir L'Homme et Marshman (2006) pour une revue détaillée sur cette question. Sur des relations spécifiques voir notamment Borillo (1996) ou Hearst (1992) sur l'hyponymie/hyperonymie, Condamines (2000) ou Otman (1996) sur la méronymie, Barrière (2001) ou Marshman (2006) sur la relation cause/effet, Hamon (2000) ou Suarez et Cabré (2002) sur la synonymie.

Néanmoins, une différence notable existe entre les marqueurs de relations conceptuelles et les marqueurs d'évolution, ces derniers ne permettant pas d'établir de lien conceptuel entre plusieurs termes/concepts. Plus précisément, le principe des marqueurs de relations est de mettre au jour des patrons dans lesquels entrent en jeu une relation conceptuelle et les termes/concepts pris dans cette relation. Par exemple, le patron classique [X est un type de Y] indique une relation d'hyperonymie/hyponymie entre les termes/concepts X et Y. Dans un texte traitant de botanique, un contexte tel que *la violette blanche et la violette des collines sont deux types de fleurs répandus en Haute-Garonne* permet d'établir un lien conceptuel entre le terme/concept hyperonyme *fleur* et les termes/concepts hyponymes *violette blanche* et *violette des collines*⁸⁰. Dans le cas des marqueurs d'évolution par contre, l'information repérable est beaucoup plus disparate et ne vise pas à établir de lien entre termes/concepts. Il ne s'agit donc pas de mettre au jour des patrons qui contiendraient un marqueur d'évolution et des termes concernés par cette évolution, mais plutôt de localiser des informations pertinentes concernant l'évolution des connaissances d'un domaine, que celles-ci impliquent des termes/concepts spécifiques ou non. Pour illustrer la spécificité des marqueurs envisageables pour repérer des informations sur l'évolution, nous donnons dans le paragraphe suivant trois exemples de travaux d'extraction d'informations évolutives.

5.1.2 Exemples d'extraction d'informations évolutives dans les textes

La démarche d'extraction de marqueurs pour repérer des informations pertinentes dans les textes est courante dans les tâches de fouille de textes et d'extraction d'informations (Ibekwe-Sanjuan, 2007). Aujourd'hui cependant, l'extraction d'informations liées à l'évolution est un point de vue encore peu répandu dans ce domaine. Il peut toutefois être illustré à travers trois approches : la veille scientifique et technologique, la gestion des risques et l'aide à la mise à jour de documents.

⁸⁰ Bien qu'il puisse arriver que l'un des termes/concepts concerné par la relation sémantique ne soit pas directement présent dans le patron recherché en corpus (par exemple dans les cas d'ellipses, d'anaphores, etc. (Marshman, 2006 : 153sq.)).

Le premier point de vue existant pour l'extraction d'informations évolutives dans les textes est celui de la veille scientifique et technologique, illustré entre autres par les travaux d'Ibekwe-Sanjuan (2005). Les principaux objectifs de la veille scientifique et technologique sont de collecter et d'organiser des informations pertinentes pour repérer – voire anticiper – les évolutions en cours dans un domaine donné, les tendances dans ce domaine, etc. L'extraction d'informations à partir de textes constitue donc une piste intéressante pour atteindre cet objectif. Dans ce cadre, Ibekwe-Sanjuan travaille à partir d'articles scientifiques rédigés en anglais et cherche à extraire dans les textes des informations sur la nouveauté et les avancées récentes dans les sciences expérimentales. Plus précisément, l'objectif de l'auteure est de proposer « un processus de guidage ou de contrôle de l'attention afin que l'expert, dont le temps est compté, puisse diriger son attention vers ce qui est réellement nouveau ou changeant » (*ibidem* : 261). L'auteure base son approche sur la recherche d'« indices langagiers de nouveauté pour détecter ce qui change dans le contenu des textes eux-mêmes » et utilise en particulier des indices⁸¹ rhétoriques qui permettent d'annoncer la présence de certains types d'informations tels que la contribution des auteurs, la présentation de résultats, les objectifs de l'article, etc. (Tableau 5.1.).

⁸¹ L'auteure choisit de parler d'*indice*, terme courant dans les tâches d'extraction d'information et de fouille de textes, alors que *marqueur* semble privilégié dans les approches terminologiques. Cette différence terminologique est assez difficile à cerner mais peut être expliquée en partie par les points de vue différents de ces approches, ainsi que par la différence de statuts et de types d'informations recherchés dans les textes, comme nous le détaillons dans cette section. Dans notre recherche, nous conservons cependant le terme *marqueur* et réservons le terme *indice* aux quatre indices détaillés dans ce manuscrit (les empreintes de fréquence, les contextes riches en connaissances évolutives, les variantes et les dépendances syntaxiques).

Type d'information	Exemples
Nouveauté	<i>new</i> <i>Here, we propose a novel [...] approach</i> <i>We discuss recent developments</i> <i>We apply this novel methodology</i> <i>If true, these claims have profound implications</i>
Résultats/Contribution/Conclusion	<i>In this paper, we show that</i> <i>Our research suggests that</i> <i>It is shown here for the first time that</i> <i>This approach may represent a step forward towards</i>
Objectifs	<i>In this [article, paper, study, research, work] we [examine, investigate, describe, present, outline]</i> <i>We discuss recent developments</i> <i>In this paper we define</i> <i>We review</i> <i>Motivation :</i>

Tableau 5.1 – Exemples de marqueurs (Ibekwe-Sanjuan, 2005 : 266)

Les indices proposés dans son étude sont sous-divisés en deux classes : des « indices textuels rhétoriques (niveau générique) » et des indices textuels « de bas niveau » (*ibidem* : 265). La première classe renvoie aux segments textuels de « division rhétorique » (*ibid.*), c'est-à-dire des indices explicites de la structure des articles. La seconde classe renvoie aux indices présents dans les phrases et les paragraphes et qui ne fournissent pas explicitement des informations sur la structure du texte. Dans l'exemple suivant (*ibid.*), sont mis en gras ce que l'auteure appelle les « indices textuels rhétoriques » et en gras souligné ce qu'elle appelle des « indices de bas niveau » :

- (i) ***Motivation*** : Introns in tRNAs are suspected to play many roles. Analysis of the bulge-helix-bulge structural motifs at the intron-exon boundaries provides some insights. The splice-sites on these structural motifs suggest that a single introncontaining-tRNA-gene can give rise to more than one tRNA product.
- Results*** : Partially overlapped mitochondrial tRNA genes are known to exist. But cytoplasmic tRNA genes overlap in archaeal methanogens in ways that are different. We present here bioinformatic evidence that in these methanogens the domain of overlap is far wider [...]

Cette répartition reste assez sommaire, mais illustre toutefois le fait que des types variés d'indices sont envisageables pour repérer des informations sur la nouveauté, indices qui

peuvent varier tant du point de vue de la classe syntaxique (adjectifs (*new*), noms (*motivation*), etc.), que du type d'information que l'on peut leur associer (nouveau, contribution, etc.), etc. Sur ce dernier point, l'auteure souligne d'ailleurs qu'il n'est pas toujours aisé de classer les indices dans une seule catégorie d'informations ; ceux qu'elle classe finalement comme « contributions » pourraient souvent être classés dans la catégorie « nouveauté » (*ibidem* : 266). Le choix des indices n'est donc pas une étape facile et doit faire l'objet d'une sélection minutieuse.

Un autre élément important dans l'approche d'Ibekwe-Sanjuan est que les éléments ciblés par les marqueurs ne constituent pas tous directement l'information recherchée mais des portions de textes qui annoncent et aident à localiser la présence d'informations pertinentes. En effet, l'auteure repère ici les zones contenant potentiellement des résultats, ou encore des zones où les objectifs sont annoncés, et non pas ces résultats ou ces objectifs eux-mêmes. Ceci constitue donc une différence importante entre la perspective évolutive et l'extraction de relations conceptuelles, où les marqueurs recherchés constituent à la fois l'indice de la présence d'une information pertinente et l'information elle-même. Nous précisons cette idée en décrivant les travaux de Sándor (2006) qui propose un autre type de recherche en veille scientifique, concernant à la fois la détection de nouveautés et d'informations relatives à la notion de risque.

Le repérage d'informations liées au risque doit être mis en lien avec l'évolution et la veille scientifique dans la mesure où il s'agit d'extraire des informations sur des types de risques à suivre et/ou à résoudre dans le temps. Pour cette tâche, Sándor (2006) propose d'extraire des traces de commentaires des auteurs dans les textes exploités, pour repérer le risque dans les domaines de la biologie et de l'alimentaire. Le Tableau 5.2 illustre certains types de marqueurs établis par l'auteure pour extraire ces informations.

Type d'information	Exemples
Les découvertes récentes contredisent les travaux existants	<i>In contrast with previous hypotheses... ... are challenging the classical role of... ...unorthodox view resolves apparent paradoxes</i>
Des matériaux dans les aliments entraînent un risque pour l'homme	<i>...risk ... for people who cause adverse health effects in humans...</i>
Citation de travaux de référence	<i>... [3, 4] is a method that can be applied to... ... has been argued frequently in the past [7, 14]...</i>

Tableau 5.2 – Exemples de marqueurs (Sándor, 2006)

L'originalité des travaux de Sándor pour la recherche d'informations sur les risques et l'évolution est qu'elle cherche explicitement à repérer des informations qui introduisent la présence d'informations pertinentes sur les risques dans les textes, plutôt que le repérage des évolutions ou des risques eux-mêmes. L'objectif de l'auteure est d'affiner le repérage d'informations pertinentes à l'aide de ces informations dites *périphériques*. Plus précisément, l'auteure insiste sur le double statut des éléments que l'on peut extraire dans les textes et présente une distinction importante : la distinction entre *information centrale* (« core information ») et *information périphérique* (« peripheral information »). Cette distinction repose sur l'observation du fait que l'extraction d'information pour la veille peut se baser :

- soit sur des éléments qui constituent eux-mêmes les informations pertinentes ; par exemple pour la veille, les informations centrales sont généralement les personnes, les événements, les termes, les dates, les relations entre les personnes, etc. (*ibidem* : 2) ;
- soit par exemple sur des commentaires d'auteurs qui indiquent, introduisent ou annoncent explicitement la présence de risques ou la présence d'autres informations pertinentes ; il s'agit alors d'informations périphériques.

Dans son article, l'auteure illustre cette distinction par l'exemple suivant (où l'information centrale est soulignée et l'information périphérique est en gras) :

- (ii) *In contrast with previous hypotheses, compact plaques form before significant deposition of diffuse A beta, suggesting that different mechanisms are involved in the deposition of diffuse amyloid and the aggregation into plaques.*

Elle explique que pour la détection des risques, *in contrast with previous hypothesis* n'est pas directement l'information recherchée, puisque cet élément ne peut être associé directement à un type de risque spécifique, mais il permet d'accéder de manière plus fiable à des contextes qui contiennent une information centrale à propos des risques.

La distinction entre ces deux types d'informations n'est pas toujours facile à établir et l'auteure elle-même ne donne pas d'arguments pour les distinguer clairement. Le seul élément mis en avant est que les informations périphériques correspondent à des commentaires de la part des auteurs sur les risques eux-mêmes et donc sur la conscience des risques qu'ils peuvent exprimer (« Likewise, if the writer is aware that a danger constitutes risk for public health, he will put forward this knowledge in some way. » (*ibidem* : 2)). Bien que cela puisse constituer une piste pour apprendre à distinguer ces deux types d'informations, il serait souhaitable d'affiner ces éléments et voir par exemple comment définir clairement ce qui relève des commentaires subjectifs des auteurs d'un autre type d'information (par exemple sur des bases linguistiques telles que la présence de certains pronoms, verbes, etc.).

Comme on le voit à travers les travaux de Sándor (2006), la diversité des marqueurs envisageables et des types d'informations que l'on peut leur associer, est également un atout pour une perspective telle que la nôtre, bien que cela implique une sélection raisonnée et *a priori*. La combinaison de marqueurs différents peut donc être vue comme un atout et une piste intéressante pour repérer des informations de manière plus précise. C'est l'optique choisie par exemple par Laignelet (2009).

Dans sa thèse (2009), l'auteure propose la mise en place d'un environnement logiciel pour maintenir à jour les fiches encyclopédiques. Ces fiches doivent en effet être mises à jour régulièrement dans la mesure où les informations qu'elles contiennent renvoient à des états de connaissances contemporains du moment de leur rédaction et susceptibles de devenir obsolètes. L'auteure cherche donc à repérer tout segment d'information évolutive, c'est-à-

dire tout « segment textuel susceptible de contenir une ou plusieurs informations qui présente(nt) la particularité de pouvoir évoluer dans le temps et/ou qui relativement à des besoins éditoriaux nécessiterai(en)t d'être réactualisé(es) » (Laignelet, 2007 : 450). Par exemple le segment mis en évidence en gras ci-dessus

- (iii) [...] *On peut ainsi savoir qu'une personne est infectée longtemps avant que la maladie ne se déclare. **Il n'existe pas à l'heure actuelle de vaccin contre le sida** [...]* (Source : corpus ATLAS (fiche Médecine – Le Sida))

devra certainement être mis à jour dans les prochaines révisions de fiches. Les travaux de Laignelet se concentrent donc sur un aspect principal d'évolution : l'obsolescence d'informations. L'originalité de ces travaux réside dans la prise en compte de marqueurs variés pour le repérage (entités nommées, marqueurs lexicaux, morphologiques, etc.) (Tableau 5.3) et notamment de marqueurs discursifs tels que les cadres de discours, les marqueurs typo-dispositionnels, ou encore les chaînes anaphoriques.

Catégories de marqueurs	Exemples
Adverbiaux temporels	<i>depuis le 16 mai 2000</i>
SN temporels	<i>les prochaines décennies</i>
Adverbes	<i>aujourd'hui</i>
Temps verbaux	<i>et contribueront à l'élaboration des premiers missiles</i>
Marqueurs aspectuo-verbaux	<i>sont en cours</i>
Sigles, noms propres, entités nommées	SIDA

Tableau 5.3 – Exemples de marqueurs (Laignelet, 2009)

Le choix de ces marqueurs différents est lié à l'objectif de l'auteure de faire émerger des configurations, c'est-à-dire des combinaisons d'indices récurrentes pour extraire des segments obsolètes des fiches encyclopédiques⁸². C'est donc la prise en compte de plusieurs marqueurs ensemble qui permet à Laignelet d'extraire des segments obsolètes dans les fiches et non pas la recherche de segments isolés. On retrouve donc l'idée des contextes riches en connaissances, ici pour des connaissances obsolètes.

⁸² La démarche de l'auteure s'inscrit dans la lignée des travaux de Péry-Woodley et Ho-Dac (2008 ; 2000) sur l'analyse du discours qui sont parmi les premières à montrer l'importance de la combinaison des indices pour l'analyse.

Bien qu'ils aient tous pour objectif d'extraire des informations sur l'évolution des connaissances à l'aide de marqueurs (ou indices) textuels, les travaux présentés dans cette section relèvent d'horizons divers, ce qui se manifeste sur plusieurs plans dont :

- leurs objectifs : veille *vs.* mise à jour de fiches ;
- les types d'évolutions en jeu : la nouveauté, le changement ou l'obsolescence ;
- les types d'informations recherchées : des portions de phrases pour Ibekwe-Sanjuan ou Sándor et des segments entiers pour Laignelet, mais également des informations centrales *vs.* des informations périphériques ;
- les types de marqueurs en jeu : des adjectifs, des noms, des syntagmes, des portions de phrases, des marqueurs discursifs, mais aussi des marqueurs isolés *vs.* des configurations.

Ces trois exemples d'approches, malgré leur hétérogénéité, offrent des éléments de réflexion centraux pour construire notre propre recherche.

Le premier de ces éléments est lié à la diversité des marqueurs envisageables pour repérer l'évolution. Cette diversité montre qu'il peut être difficile de saisir et de caractériser les marqueurs. Du point de vue des types d'informations recherchés pour notre recherche, ceci est d'autant plus notable dans notre étude qu'il s'agit d'une approche exploratoire de l'évolution. La sélection des marqueurs faite *a priori* doit donc être claire et prendre en compte au mieux la diversité des aspects d'évolution que l'on peut s'attendre à rencontrer. Cette sélection doit de plus tenir compte des genres de textes dans lesquels l'information est recherchée. Par exemple, les commentaires d'auteurs ou les indices rhétoriques considérés par Ibekwe-Sanjuan ne sont prévisibles que dans la mesure où les textes choisis sont des articles scientifiques dont on peut faire l'hypothèse qu'ils contiennent ce type d'information. Il est très peu probable en effet de rencontrer une information de type « contribution » dans des genres de textes autres qu'argumentatifs. Il faut donc prendre en compte cet élément pour définir des marqueurs adaptés (voir notamment (Condamines, 2003b) sur l'impact du genre pour la définition de marqueurs). Du point de vue de la nature des marqueurs à sélectionner, il est également possible d'envisager de nombreuses pistes telles que des

marqueurs discursifs, morphologiques, structurels, lexicaux, etc. Dans le cas d'une démarche exploratoire telle que la nôtre, il est donc encore une fois primordial de savoir circonscrire la liste des marqueurs en fonction de ces différentes possibilités.

Le second élément est lié au fait que les marqueurs peuvent être combinés pour extraire des contextes riches en connaissances d'évolution. Il nous semble qu'il s'agit d'une caractéristique centrale pour notre étude. En effet, nous l'avons vu, les marqueurs d'évolution possibles ne peuvent pas se construire sous la forme de patrons qui mettraient en relation un type d'évolution et des termes/concepts pris dans cette évolution ; les marqueurs d'évolution aident plutôt à cibler des contextes qui contiennent des informations pertinentes. De plus, comme le montre Laignelet et comme nous le verrons plus loin dans ce chapitre (§5.3.1, p.169), la combinaison des marqueurs doit être envisagée pour mettre au jour des contextes de manière plus précise et fiable.

La proposition de combiner les marqueurs pour extraire des contextes doit également être mise en lien avec un troisième élément lié à la démarche adoptée dans notre recherche. Nous adoptons en effet dans notre recherche une démarche linguistique en corpus et en collaboration avec des experts. En d'autres termes, ce qui nous intéresse ici est le fait que ces contextes soient analysés et interprétés par l'analyste et des experts. Dans le cadre de l'extraction d'information cependant, ce point de vue sur l'interprétation est partagé par de nombreux chercheurs, notamment en terminologie (par exemple Condamines, 2003a ; Marshman, 2006 ; Meyer, 2001 ; Meyer, *et al.*, 1999). Comme le souligne d'ailleurs Marshman (2006 : 3) :

« human interpreters of language will always be able to provide a more informed interpretation of that language than even the most highly developed automatic application, and can thus identify and correct many problems in automatic analysis. Moreover, human users are the best judges of the complex extra-linguistic factors surrounding the results produced by an automatic tool, which in large part determine the pertinence of any piece of information in a use situation. »

Nous partageons ce point de vue et nous inscrivons dans une démarche de repérage automatisé de marqueurs et de contextes riches en connaissances évolutives associée à une

analyse « humaine ». Ceci est d'autant plus important que, d'un point de vue méthodologique, l'observation de contextes d'évolution peut aussi être considérée comme une première étape de l'observation de l'évolution en corpus dans laquelle l'analyste est confronté directement aux corpus. L'extraction d'informations globales telle qu'elle est envisagée ici lui permet donc de s' « immerger » dans les textes et lui offre une présentation riche et globale d'aspects de l'évolution détaillés en corpus. Dans la mesure où l'analyste (linguiste ou terminologue) est rarement un expert ou un historien du domaine, l'analyse manuelle de ces contextes offre un angle d'approche intéressant pour aborder la question de l'évolution.

5.2 Extraction des contextes d'évolution à l'aide de marqueurs

5.2.1 Marqueurs définis pour cette recherche

L'extraction de contextes riches en connaissances évolutives passe dans un premier temps par la définition de marqueurs d'évolution.

Les marqueurs sont définis *a priori* et identifiés sur la base d'une démarche introspective et itérative, que l'on retrouve dans les travaux de Méla (2004) ou Leroy (2004) par exemple. Plus précisément, nous partons d'une première liste de quelques marqueurs définis sur la base de notre intuition de locuteur. Cette liste est projetée dans les corpus à l'aide de l'outil TerminoWeb (développé par Barrière (Barrière & Agbago, 2006) et décrit au Chapitre 3) qui permet ensuite de visualiser les extraits de corpus dans lesquels apparaissent les marqueurs de la liste (Figure 3.3, p.96). L'observation des résultats permet d'ajuster la première liste intuitive en y ajoutant de nouveaux marqueurs possibles qui apparaissent dans les résultats obtenus mais qui n'avaient pas été identifiés. Ils permettent également d'éliminer les marqueurs qui apportent trop de bruit. Ces allers-retours entre la liste de marqueurs et les extraits en corpus sont reproduits de manière itérative jusqu'à ce que la liste obtenue soit satisfaisante. À titre d'exemple, nous avons effectué ces allers-retours à 7 reprises.

Comme nous venons de le voir, la sélection *a priori* des marqueurs n'est pas une étape aisée et repose sur une série de choix justifiés par les besoins de l'analyse. Les deux principaux

choix concernent les catégories de marqueurs et les types d'informations recherchés. Dans notre cas, les marqueurs choisis sont des marqueurs lexico-syntaxiques (noms, adjectifs, adverbes, verbes, locutions verbales et syntagmes prépositionnels essentiellement).

Du point de vue des types d'informations, la liste finalement définie pour cette recherche contient une quarantaine de marqueurs répartis en neuf « groupes » d'interprétations qui couvrent de manière satisfaisante le champ des évolutions possibles en diachronie courte.

Nous avons défini ces groupes sur la base de trois observations :

1. l'étude de l'évolution peut se baser sur la datation d'événements ;
2. les évolutions typiquement prises en considération sont la nouveauté et la disparition de termes/concepts ;
3. les évolutions scientifiques et techniques sont souvent motivées par l'apparition de nouveaux besoins dans le domaine, qui entraînent alors progrès et améliorations.

À partir de ces trois observations les marqueurs proposés sont répartis comme suit :

1. Datation d'événements
 - a. Marqueurs de datation
 - b. Marqueurs d'ancrage dans le présent/l'actualité
 - c. Marqueurs d'ancrage dans le passé
2. Nouveauté et disparition de termes ou de concepts
 - a. Marqueurs de nouveauté ou d'apparition d'une entité
 - b. Marqueurs de disparition, d'obsolescence
3. Nouveaux besoins potentiels
 - a. Marqueurs d'insuffisance
 - b. Marqueurs de comparaison
 - c. Marqueurs de remplacement ou de succession
 - d. Marqueurs d'amélioration ou de développement
 - e. Marqueurs de prédiction (futur)

Les paragraphes suivants détaillent l'ensemble des marqueurs retenus en fonction des neuf groupes que nous proposons.

1.a. Datation d'événements : marqueurs de datation

Ces marqueurs correspondent à la datation des événements liés à des concepts du domaine, et permettent d'établir des bases dans la chronologie de ce domaine. Ils apportent des informations riches pour commencer une exploration et poser des étapes dans l'histoire de certains termes/concepts.

3 marqueurs proposés ⁸³ :

- ***an(s) | année(s) [date]***
 - (i) *Les polynômes de Zernike ont pris de l'importance quand les ordinateurs ont permis de les calculer rapidement (ils ont été inventés dans le début des années 40 ! mais jamais exploités)*
- ***en [date]***
 - (ii) *Le premier terminal est monté sur Spot 4 (lancé en 2000), le second sur ARTEMIS (lancé en 2001)*
- ***depuis quelques années***
 - (iii) *Depuis quelques années, la technique d'usinage ionique sans contact avec le miroir est opérationnelle*

1.b. Datation d'événements : marqueurs d'ancrage dans le présent/l'actualité

Il s'agit de marqueurs qui ancrent les informations dans le présent et indiquent que les informations données sont d'actualité au moment de la rédaction des textes. Ce type de marqueurs permet de faire le point – du moins partiellement – sur l'état de connaissances du domaine à un moment précis (celui de la rédaction du corpus).

7 marqueurs proposés :

- ***à ce jour***
 - (iv) *on a pu lire que les plus grandes barrettes à ce jour comprennent autour de 12 000 détecteurs*
- ***actuellement***
 - (v) *Ces lidars sont actuellement extrêmement demandés par la communauté scientifique*

⁸³ Dans les extraits proposés, plusieurs marqueurs peuvent apparaître. Par souci de clarté et à des fins d'illustration nous ne soulignons cependant que le marqueur abordé.

- ***aujourd'hui***
 - (vi) le Zérodur, le plus mature, et le Carbure de Silicium (SiC) fritté qui sont utilisés aujourd'hui dans la majorité des cas
- ***classique***
 - (vii) Baffle semi-spéculaire Baffle semi- spéculaire : La structure est celle d'un baffle classique
- ***disponible***
 - (viii) Les outils et les données technologiques disponibles permettent de limiter les risques lors du développement d'une nouvelle fonction
- ***exister/existant***
 - (ix) De nombreuses variantes de ce télescope existent
- ***maintenant***
 - (x) Le signal est stocké dans une mémoire (bande magnétique et maintenant, mémoire à état solide)

1.c. Datation d'événements : marqueurs d'ancrage dans le passé

Il s'agit de marqueurs qui ancrent les connaissances dont il est question dans le passé. Ils permettent de relever certaines informations valides à un temps antérieur à celui de la rédaction du corpus.

3 marqueurs proposés :

- ***autrefois***
 - (xi) Autrefois, la technologie des détecteurs ne permettait pas d'avoir des barrettes ou des mosaïques (CCD)
- ***histoire***
 - (xii) soit au tout début de l'histoire des lasers
- ***passé***
 - (xiii) Jusqu'à un récent passé, ces instruments n'étaient pas disponibles à cause de leur difficulté de polissage

2.a. Nouveauté et disparition de termes/concepts : marqueurs de nouveauté et apparition

Les marqueurs de ce groupe annoncent la présence de termes/concepts présentés comme nouveaux ou récents. Comme l'illustre l'exemple du marqueur *inventer*, l'état de nouveauté

n'est pas toujours relatif au moment de la rédaction du corpus, mais peut être relatif à une date précise spécifiée dans le contexte (dans l'exemple (xvi), dans les années 40).

5 marqueurs proposés :

- ***apparaître***
(xiv) *Partant de ces produits sont apparues des applications pour les caméras d'observation de la Terre*
- ***arrivée***
(xv) *observation ou imagerie spectrale à haute résolution, surveillance de l'environnement, l'arrivée du multimédia, la télévision numérique à haute définition*
- ***inventer***
(xvi) *ils ont été inventés dans le début des années 40*
- ***nouveau***
(xvii) *Un produit nouveau est apparu depuis quelques années sur le marché, il s'agit de multi barrettes*
- ***récent***
(xviii) *CCD aminci : Le substrat dans lequel est réalisé le détecteur est aminci. [...] Cette technologie, récente, est très appréciée*

2.b. Nouveauté et disparition de termes/concepts : passé, disparition, obsolescence

Dans cette classe, les marqueurs considérés indiquent la présence de termes/concepts qui ont été abandonnés ou sont devenus obsolètes. Notons que dans nos corpus ceux-ci s'avèrent très rares.

1 marqueur proposé :

- ***abandonner***
(xix) *Ce type d'architecture a été étudié en phase 0 du projet Pleiades pour être finalement abandonnée pour des problèmes d'obscurité*

3.a. Nouveaux besoins potentiels : marqueurs d'insuffisance

Les marqueurs considérés dans cette classe ciblent les lacunes et les insuffisances de certains concepts du domaine. De fait, l'hypothèse est que ces concepts seront potentiellement liés à

des progrès à venir pour pallier ces lacunes et répondre aux besoins du domaine. Il s'agit donc de concepts dont l'évolution est à suivre dans le temps.

4 marqueurs proposés :

- **malheureusement**
(xx) *Malheureusement, ces hybrides nécessitent des photocathodes d'efficacité quantique limitée (10 à 20 % maxi) et ne travaillent que dans le visible*
- **pas assez**
(xxi) *les mesures ne sont pas assez précises et on ne peut pas obtenir toutes les altitudes et observer partout*
- **pas encore**
(xxii) *Comme la technologie des diodes laser ne permet pas encore de monter trop haut en puissance [...] on juxtapose sur le même télescope plusieurs diodes laser [...]*
- **pas satisfaisant**
(xxiii) *la qualité des filtres [...] employés pour la reprographie ne sont pas satisfaisants pour le spatial, et il faut envisager des filtres plus spécifiques*

3.b. Nouveaux besoins potentiels : marqueurs de comparaison

Les marqueurs recensés dans cette classe expriment une relation de comparaison entre au moins deux concepts. Il peut s'agir d'une comparaison entre deux concepts différents, mais également entre les différents états ou « générations » d'un seul et même concept qui a évolué. Les hypothèses d'évolution associées peuvent être d'ordres divers : apparition d'une nouvelle technique ou d'un nouveau concept qui remplace potentiellement une technique (un concept) moins performant(e), progrès/améliorations entre différentes générations d'un même concept, amélioration des propriétés de certains concepts, etc.

3 marqueurs proposés :

- **meilleur (que)**
(xxiv) *Le CNET a réalisé des lasers à puits quantiques dont le courant de seuil est inférieur à 2 et le gain différentiel de l'ordre de 0,3 W/A (trois fois meilleur qu'un laser classique)*
- **moins [...] (que)**
(xxv) *un modèle macroscopique simplifié, et de ce fait moins précis ou avec un domaine de validité plus restreint, mais plus rapide en terme de simulation*

- **plus [...] (que)**
(xxvi) *On cite, par exemple les nouveaux matériaux pour la fabrication de miroirs, les allègements plus poussés, les plans focaux plus "intégrés"*

3.c. Nouveaux besoins potentiels : marqueurs de remplacement/succession

Dans cette classe, les marqueurs permettent de repérer les cas où un concept fait ou a fait place à un autre. Le « remplaçant » est susceptible de devenir la nouvelle référence dans le domaine alors que le « remplacé » est susceptible de devenir obsolète. Il peut être important de suivre ces deux termes/concepts dans le temps.

2 marqueurs proposés :

- **remplacer**
(xxvii) *les détecteurs CMOS vont remplacer les CCD*
- **successeur**
(xxviii) *Polder, son successeur Parasol, ou encore Modis (USA) [...], sont d'autres exemples*

3.d. Nouveaux besoins potentiels : marqueurs d'amélioration/développement

Les marqueurs regroupés dans cette classe permettent d'extraire trois types d'informations proches sur l'évolution :

- des informations sur des concepts améliorés dans le domaine,
- des informations sur des concepts en cours de développement, en progrès, ou susceptibles de faire l'objet de travaux,
- des informations sur des concepts en essor ou en développement au moment de la rédaction.

9 marqueurs proposés :

- **être à l'étude**
(xxix) *Comme il existe d'autres contraintes, il faut utiliser d'autres "astuces" pour parfaire cette réjection [...] et même d'autres principes instrumentaux sont à l'étude*

- **amélioration/améliorer**
(xxx) Les améliorations technologiques portent généralement sur les techniques de confinement de la lumière et des porteurs d'électrons (électrons et trous)
- **développement/développer**
(xxxi) Les instruments développés seront donc des caméras d'imagerie à haute résolution spatiale
- **encourageant**
(xxxii) A la vue des résultats très encourageants obtenus
- **essor**
(xxxiii) L'optique en plein essor : les techniques "SOL" actuelles
- **évoluer/évolution**
(xxxiv) Les lasers solides ont, eux, évolué vers la production de lasers compacts
- **progrès**
(xxxv) On notera les progrès importants accomplis dans la fabrication de barrettes et mosaïques CCD infrarouges
- **prototype**
(xxxvi) on voit apparaître des prototypes de deux miroirs
- **réduire**
(xxxvii) Ceci conduit à réduire les dimensions des antennes et certains équipements et composants

3.e. Nouveaux besoins potentiels : marqueurs de prédiction

Ces marqueurs permettent de repérer des informations sur des concepts dont les experts/rédacteurs pensent qu'ils seront centraux dans l'avenir. Les concepts concernés devront également être suivis dans le temps pour confirmer ou infirmer les prédictions à leur sujet.

5 marqueurs proposés :

- **avenir**
(xxxviii) On peut facilement imaginer que l'avenir sera dans des instruments plus "simples", miniaturisés au maximum

- ***envisager***
(xxxix) *Dans le plus proche avenir, il faudra envisager sérieusement des systèmes de télécommunication à très hauts débits*
- ***futur***
(xli) *Mais il est certain que leur évolution est à suivre de près pour les futures applications spatiales*
- ***imaginer***
(xlii) *On pourra aussi imaginer d'autres utilisations des lasers semi-conducteurs pour les communications intra-satellite ou pour la métrologie à bord*
- ***prometteur***
(xliii) *ces instruments (détection directe) semblent très prometteurs*

Ces différents groupes de marqueurs nous incitent à faire écho à la distinction entre information centrale et information périphérique. En effet, la liste que nous proposons illustre la difficulté que l'on peut avoir à statuer sur le type d'information en jeu. Prenons l'exemple d'un marqueur tel que *aujourd'hui* :

- (xliiii) *Aujourd'hui des centaines de millions de transistors sont implantés sur une plaquette avec des technologies forcément de plus en plus fines ; on est passé en dessous du 1 μm . pour atteindre 0,8 μm et on travaille déjà sur la technologie à 0,5 μm .*

Dans ce premier exemple (xliiii), le marqueur n'est pas un commentaire de la part du rédacteur sur l'évolution du domaine (comme Sándor définissait les informations périphériques), ni une information centrale recherchée. Cependant il est pertinent pour repérer des informations sur l'évolution des connaissances puisqu'il permet de « dater » les assertions qui suivent et de fournir ainsi des informations sur l'état des connaissances du domaine au moment de la rédaction des textes. Il s'agit donc d'autres informations périphériques qui ne sont pas des commentaires d'auteurs.

Dans l'exemple suivant :

- (xliv) *Cependant dans ce cas, la qualité des filtres, tant vis-à-vis du matériau que de leurs performances, employés pour la reprographie ne sont pas satisfaisants pour le spatial.*

le marqueur en gras, *pas satisfaisant*, peut être considéré comme un commentaire de l’auteur puisque ce dernier exprime clairement le fait que les filtres en question présentent des limites. Dans ce cas, on pourrait dire qu’il s’agit d’une information périphérique. Or, on peut également juger que cette information est centrale pour comprendre l’évolution du domaine puisqu’elle présente une information sur la qualité et les performances des filtres au moment de la rédaction. De la même manière, dans l’exemple (xlv), *nouveau* est à la fois une information périphérique et une information centrale : il indique la présence d’éléments pertinents d’un point de vue évolutif (le fait que les *instruments infrarouges* sont concernés par une évolution), ainsi qu’une information centrale dans la mesure où ce marqueur qualifie l’évolution en jeu (il s’agit de nouveauté).

- (xlv) *Simplicité d’emploi et économie d’un système de refroidissement sont les deux avantages principaux de ces nouveaux instruments infrarouges.*

Ces quelques exemples visent à montrer que tous les marqueurs n’ont en effet pas nécessairement le même statut mais qu’il peut être difficile de les caractériser clairement.

Ceci permet de mettre en avant le fait que les différents types d’informations proposés dans nos neuf groupes sont donc pour la plupart complémentaires. Nous en donnerons d’autres illustrations dans la suite de ce chapitre, mais comme nous l’avons souligné dans l’exemple (xvi) (p.162) il est clair notamment que les marqueurs de datation sont souvent complémentaires des autres types de marqueurs puisqu’ils permettent de nuancer et d’interpréter correctement des informations sur la nouveauté, l’obsolescence, etc. (dans l’exemple (xvi), le marqueur de datation *dans les années 40* permet de compléter l’information donnée par le marqueur *inventés*). Ceci confirme donc l’importance de combiner les marqueurs pour extraire des contextes pertinents ainsi que l’importance de l’interprétation « humaine » pour l’analyse.

Le paragraphe suivant décrit la manière dont nous extrayons les contextes riches en connaissances évolutives à partir des marqueurs que nous venons de décrire.

5.2.2 Extraction de contextes riches en connaissances évolutives

Comme nous l'avons dit, l'analyse proposée ne se restreint pas à l'observation de marqueurs mais vise à extraire des contextes riches en connaissances évolutives à l'aide de ces marqueurs. Pour cela, les contextes recherchés ici sont définis comme des portions de textes qui contiennent au moins un marqueur d'évolution et un ou plusieurs terme(s)/concept(s) du domaine.

L'extraction est menée encore une fois à l'aide de l'outil TerminoWeb, ainsi que de programmes Perl écrits par nos soins.

À partir de cette définition des contextes recherchés, la principale difficulté est de savoir comment les délimiter. En effet, nous l'avons dit, les marqueurs d'évolution ont ceci de particulier qu'ils ne peuvent généralement pas être construits sous la forme de patrons qui les associeraient à des termes/concepts du domaine. Dans l'approche choisie ici, l'objectif est plutôt de mettre au jour des « zones » (ou des segments à explorer, pour reprendre la terminologie de Laignelet (2009)) que de véritables contextes complets et clos.

La délimitation des contextes se fait le plus souvent en considérant le niveau de la phrase (voir notamment Ibekwe-Sanjuan (2005)). On pourrait considérer alors qu'un contexte riche en connaissances évolutives est une phrase qui contient au moins un marqueur et un terme/concept du domaine. Mais en observant manuellement les résultats obtenus dans nos corpus, nous avons finalement porté notre choix de délimitation sur le paragraphe. En effet, le contexte de la phrase s'est avéré souvent trop court pour extraire des informations pertinentes complètes ; le paragraphe semble de fait un bon compromis entre la phrase et une section titrée, souvent trop longue et trop hétérogène du point de vue de l'information contenue. Néanmoins, certains paragraphes peuvent ne pas contenir toute l'information pertinente. C'est le cas notamment des paragraphes qui reprennent un ensemble d'informations décrites dans les paragraphes précédents. Par exemple, dans le contexte d'évolution suivant (où les marqueurs sont soulignés et les termes/concepts en gras) :

- (xlvi) Ces nouvelles technologies n'ont pas encore réellement fait leur entrée dans le spatial, car elles sont plus récentes que celle des **semi-conducteurs**. Elles ont donc encore à prouver leurs potentialités dans ce domaine. Mais il est certain que leur évolution est à suivre de près pour

*les futures applications spatiales, où faibles **encombrement** et **masse**, faible **consommation électrique** et bas coût seront des caractéristiques déterminantes.*

l'anaphore en début de paragraphe (*ces nouvelles technologies*) indique que la liste des technologies marquées comme nouvelles ne peut être retrouvée que plus haut dans le texte, dans les paragraphes précédents. Dans ce cas, et avec la méthodologie mise en place dans cette étude, le linguiste/terminologue devra explorer (manuellement) un contexte un peu plus large pour récolter l'ensemble des informations pertinentes. Ces cas sont cependant assez peu fréquents.

Notons que d'autres démarches pour la délimitation de contextes sont possibles, par exemple à l'aide de calculs statistiques, sur le modèle du TextTiling de Hearst (1997)⁸⁴.

Les contextes riches en connaissances évolutives sur lesquels nous basons nos observations sont donc des paragraphes dans lesquels au moins un marqueur et un terme/concept sont présents⁸⁵. Nous décrivons les analyses que ces contextes permettent de construire et montrons notamment en quoi la combinaison de plusieurs marqueurs influe sur la pertinence des contextes extraits.

5.3 Observations en corpus : Projection des termes en contextes

Une fois les contextes extraits, l'objectif est de chercher à observer la liste des termes du domaine y gravitant effectivement. Deux points de vue se complètent pour analyser les contextes :

- analyser chacun des contextes,
- ou observer la répartition des termes dans ces contextes.

5.3.1 Évaluation de la pertinence des contextes et des marqueurs

La sélection proposée ici est une proposition de marqueurs susceptibles d'indiquer des contextes pertinents pour repérer l'évolution des connaissances du domaine spatial. Il est

⁸⁴ C'est ce qu'ont tenté notamment Laignelet et Pimm (2007).

⁸⁵ Obtenue ici avec Syntex.

important cependant de noter que plusieurs éléments influencent ce choix, dont le type des textes qui composent les corpus influence grandement cette sélection. En effet, mais cela a été souligné dans le cas de l'extraction de marqueurs de relation (par Borillo, 1996 ; Condamines, 2003a ; 2003b notamment), les textes didactiques ou qui s'adressent à des non spécialistes sont susceptibles de contenir beaucoup plus de contextes riches en connaissances d'évolution que les textes techniques qui s'adressent à des spécialistes. C'est ce que l'on remarque ici dans nos deux corpus : si ces contextes sont assez nombreux dans le corpus TTVS, ils sont extrêmement marginaux dans le corpus DORIS. De fait, les observations menées dans ce chapitre le seront exclusivement pour le corpus TTVS, et l'on devra retenir l'importance que peut avoir le genre et/ou le degré de spécialité des textes sur la définition et la mise en œuvre de cet indice.

De manière générale, les contextes extraits contiennent un bruit assez important : un peu plus de la moitié d'entre eux ne sont pas pertinents, comme l'illustre le Tableau 5.4.

	Nombre de contextes d'évolution	Nombre de contextes pertinents	Précision (%)
TTVS1994	255	108	42,35
TTVS1998	420	160	38,10
TTVS2002	527	207	39,28
Total	1202	475	39,52

Tableau 5.4 – Pertinence des contextes d'évolution

Ce bruit est principalement dû à deux phénomènes. Le premier est lié à la présence de marques d'incertitude ou de négation difficiles à anticiper et qui viennent modifier l'interprétation associable *a priori* au marqueur. En effet, la plupart des marqueurs proposés sont susceptibles d'être modifiés, niés ou « nuancés » par le rédacteur dans les textes. Ainsi, dans l'extrait

- (xlvi) *Dans le domaine strict des communications, on peut se poser la question de savoir si la détection hétérodyne supplantera la détection directe, grâce à ses grandes potentialités, ou si, au contraire, l'évolution très rapide des performances des lasers semi-conducteurs consacrera cette dernière par sa simplicité d'emploi*

l'évolution est présentée sous forme d'hypothèse. L'interprétation doit donc être nuancée voire ne pas être prise en compte.

Le second phénomène, plus fréquent, renvoie au fait que dans certains contextes, l'information temporelle extraite n'est pas liée au domaine en général mais à une situation donnée, un « contexte de discours », phénomène qui se rapporte à la notion de « relation discursive contextuelle » décrite par Borillo (1996 : 122). L'auteure propose cette notion dans le cadre de l'exploration de la relation conceptuelle d'hyponymie/hyperonymie. Elle remarque que, dans certains cas, la relation extraite n'est pas stable et lexicale et n'intervient que dans une situation discursive donnée. Dans le cas des contextes riches en connaissances d'évolution, ceci se traduit par des cas où l'évolution en jeu ne renvoie pas à une évolution stable et avérée dans les connaissances générales du domaine, mais à une évolution dans le cadre d'une situation spécifique telle qu'une expérience scientifique. Par exemple :

(xlvi) *L'approche est d'augmenter progressivement les rapports D/d . Au début, on est à la limite de l'occultation centrale : la pupille est "presque pleine".*

Dans cet extrait, on voit que les marqueurs en gras et soulignés renvoient à des actions ou des étapes propres à une expérience décrite et ne peuvent être associés à des connaissances générales du domaine.

Ce cas de figure a été particulièrement présent pour certains marqueurs, tels que *changer/changement* qui a dû finalement être éliminé de la liste finale dans la mesure où il apportait trop de bruit. En effet, bien que certains contextes aient été pertinents :

(xlix) *Tout semble indiquer que les réseaux de communications vont être changés par l'arrivée des technologies de la photonique.*

ce marqueur prend le plus souvent dans nos corpus une valeur propre à un contexte discursif spécifique :

(l) *il ne nous reste plus qu'à implanter par diffusion des impuretés n dans le substrat pour changer la nature du semi-conducteur dans la zone délimitée par le masque.*

Malgré ces erreurs, la pertinence des contextes reste intéressante, d'autant plus qu'elle peut dépendre du nombre de marqueurs contenus par contexte. Cette observation, que l'on retrouve notamment chez Malaisé *et al.* (2005), signifie que plus un contexte contient de marqueurs, plus il est pertinent du point de vue de l'évolution. Il est donc possible de filtrer les contextes les plus pertinents en fonction du nombre de marqueurs qu'ils contiennent. Pour illustrer cet aspect, le Tableau 5.5 présente la répartition des différents contextes évolutifs en fonction du nombre de marqueurs qu'ils contiennent. Le Tableau 5.6 présente le nombre et la proportion de contextes **pertinents** en fonction du nombre de marqueurs.

		TTVS1994	TTVS1998	TTVS2002	Total
Nombre de marqueurs par contexte	8	0	1	1	2
	7	1	1	1	3
	6	0	3	4	7
	5	1	5	5	11
	4	7	6	9	22
	3	7	24	38	69
	2	54	83	120	257
	1	185	297	349	831
	Total	255	420	527	1202

Tableau 5.5 – Répartition des contextes en fonction du nombre de marqueurs

		TTVS1994		TTVS1998		TTVS2002		Total	
		Nb	%	Nb	%	Nb	%	Nb	%
Nombre de marqueurs par contexte	8	-	-	1	100	1	100	2	100
	7	1	100	1	100	1	100	3	100
	6	-	-	3	100	4	100	7	100
	5	1	100	5	100	5	100	11	100
	4	7	100	5	83,33 ⁸⁶	9	100	21	94,44
	3	5	71,43	17	70,83	27	71,05	49	71,10
	2	23	42,59	35	42,17	48	40	106	41,59
	1	71	38,38	93	31,31	112	32,09	276	33,93
	Total	108	42,35	160	38,10	207	39,28	475	39,52

Tableau 5.6 – Nombre et proportion de contextes pertinents en fonction du nombre de marqueurs

⁸⁶ Ce chiffre, relativement bas comparé aux 100% de précision obtenus, doit être lu avec précaution. En effet bien que la pertinence soit dégradée de près de 17%, dans les faits un seul contexte contenant quatre marqueurs n'est pas pertinent sur les 22 contextes concernés. Relativement au nombre de contextes concernés, il ne s'agit pas d'un bruit considérable malgré sa proportion de 17%.

Ces deux tableaux permettent de confirmer que plus les marqueurs sont nombreux dans un contexte donné plus il est probable que celui-ci soit pertinent. Ainsi, dans les cas où les contextes contiennent de 3 à 8 marqueurs (en gris dans le Tableau 5.6), de 71 à 100% des contextes sont pertinents, soit 93 contextes sur un total de 114. Par contre, les performances se dégradent assez significativement dès que le contexte ne contient qu'un ou deux marqueur(s). Cependant, il est clair également que plus les contextes contiennent de marqueurs moins ils sont nombreux.

Sur la base de cette observation, ne seront donc pris en considération dans la suite de ce chapitre que les contextes qui contiennent au moins 3 marqueurs dans la mesure où il s'agit des cas où plus de la moitié des contextes extraits sont pertinents. Le taux de précision moyen atteint alors 81,57% dans l'ensemble du corpus TTVS (Tableau 5.7).

Sous-corpus	Nombre de contextes d'évolution	Nombre de contextes pertinents	Précision
TTVS1994	16	14	87,50 %
TTVS1998	40	32	80 %
TTVS2002	58	47	81,03 %
Total	114	93	81,58 %

Tableau 5.7 – Pertinence des contextes d'évolution

À partir de ces contextes, la précision des marqueurs a été évaluée : pour chacun, la proportion d'occurrences pertinentes est calculée par rapport à l'ensemble des occurrences du marqueur dans les contextes. Par exemple, dans un contexte tel que :

- (li) Une nouvelle classe d'instruments devient nécessaire : l'interférométrie qui permet d'augmenter la résolution spatiale. Cette technique est associée à une autre, la synthèse d'ouverture qui permet de reconstruire progressivement l'image de l'objet à partir des informations de son spectre en fréquences spatiales (contraste et phase). Cette technique fait appel à l'interférométrie à grande base. Déjà implantée au sol (VLTi), pour des missions relativement simples, et grâce aux progrès récents des technologies de l'optique, l'interférométrie pourra devenir la nouvelle composante de l'instrumentation optique spatiale. (TTVS)

on compte 4 occurrences de marqueurs (en gras et soulignés), dont 2 pour le marqueur *nouvelle*. Ces 4 occurrences sont pertinentes.

Par contre, dans un contexte tel que :

- (lii) *Les exemples ne manquent pas pour convaincre le lecteur que, dans le plus proche avenir, il faudra envisager sérieusement des systèmes de télécommunication à très hauts débits et suffisamment évolutifs pour s'adapter à des besoins très particuliers.* (TTVS)

on compte 3 occurrences de marqueurs, mais seulement 2 pertinentes (*évolutifs* indique ici une propriété des systèmes de télécommunication, et non pas une information sur l'évolution du domaine). Le contexte reste cependant pertinent.

Dans un second temps, nous nous intéressons à la performance des marqueurs seuls.

Le Tableau 5.8 indique la précision des marqueurs en fonction des trois groupes principaux d'interprétation auxquels ils appartiennent (§5.2.1, p.159).

	Occurrences totales	Pertinents	Non pertinents	Précision (%)
Marqueurs de datation	91	77	14	89,01
Nouveauté et disparition de termes/concepts	68	61	7	89,86
Nouveaux besoins potentiels	320	229	91	71,47

Tableau 5.8 – Précision des trois groupes principaux de marqueurs (TTVS)

Chacun des trois groupes présente une précision tout à fait satisfaisante, supérieure à 70%. Le groupe « Nouveaux besoins potentiels » est cependant un peu moins précis que les deux autres, ce qui s'explique en partie par le fait qu'il contient beaucoup de marqueurs « multi-valeurs », c'est-à-dire les marqueurs dont l'information temporelle est susceptible de s'ancrer dans un contexte discursif spécifique et qui n'indiquent alors pas une évolution des connaissances du domaine (cf. Borillo, 1996, *supra*).

Le Tableau 5.9 détaille la précision de chacun des marqueurs dans l'ensemble du corpus TTVS, répartis dans les neuf groupes identifiés *supra* (§5.2.1, p.159).

		Occurrences totales	Pertinents	Non pertinents	Précision (%)
Marqueurs de datation	Marqueurs de datation				
	<i>années [date]</i>	6	6	0	100
	<i>en [date]</i>	2	2	0	100
	<i>depuis quelques années</i>	3	3	0	100
	Marqueurs d'ancrage dans le présent/l'actualité				
	<i>à ce jour</i>	3	3	0	100
	<i>actuellement/actuel</i>	28	28	0	100
	<i>aujourd'hui</i>	9	9	0	100
	<i>classique</i>	10	7	3	70
	<i>disponible</i>	9	6	3	66,67
	<i>exister/existant</i>	8	5	3	62,5
	<i>maintenant</i>	8	7	1	87,5
	Marqueurs d'ancrage dans le passé				
	<i>autrefois</i>	1	1	0	100
	<i>histoire</i>	2	2	0	100
	<i>passé</i>	2	2	0	100
Nouveauté et disparition de termes/concepts	Marqueurs de nouveauté/apparition				
	<i>apparaître</i>	18	15	3	83,33
	<i>arrivée</i>	4	2	2	50
	<i>inventer</i>	1	1	0	100
	<i>nouveau</i>	33	31	2	93,94
	<i>récent</i>	11	11	0	100
	Marqueurs du passé, disparition, obsolescence				
	<i>abandonner</i>	2	2	0	100
Nouveaux besoins potentiels	Marqueurs d'insuffisance				
	<i>malheureusement</i>	7	7	0	100
	<i>pas assez</i>	2	1	1	50
	<i>pas encore</i>	13	13	0	100
	<i>pas satisfaisant</i>	2	2	0	100
	Marqueurs de comparaison				
	<i>meilleur (que)</i>	24	16	8	66,67
	<i>moins [...] (que)</i>	9	5	4	55,56
	<i>plus [...] (que)</i>	76	48	28	63,16
	Marqueurs de remplacement/succession				
	<i>remplacer</i>	5	5	0	100
	<i>successeur</i>	7	7	0	100

Marqueurs d'amélioration/développement				
<i>être à l'étude</i>	1	1	0	100
<i>amélioration/améliorer</i>	35	19	16	54,29
<i>développement/développer</i>	22	17	5	77,27
<i>encourageant</i>	1	1	0	100
<i>essor</i>	2	2	0	100
<i>évoluer/évolution/évolutif</i>	23	18	5	78,26
<i>progrès</i>	12	12	0	100
<i>prototype</i>	2	2	0	100
<i>réduire</i>	43	23	20	53,49
Marqueurs de prédiction				
<i>avenir</i>	6	6	0	100
<i>envisager</i>	13	11	2	84,62
<i>futur</i>	7	7	0	100
<i>imaginer</i>	5	5	0	100
<i>prometteur</i>	2	2	0	100
TOTAL	479	373	106	77,87

Tableau 5.9 – Précision de chaque marqueur dans l'ensemble du corpus (TTVS)

Ce tableau confirme le fait que les performances obtenues grâce à la liste définie pour cette recherche sont de bonne qualité. La précision des marqueurs varie de 50% à 100%, pour une moyenne de 77,45% de précision et près de 60% de ces 42 marqueurs présentent une précision de 100%.

Pour compléter ces résultats, un tableau qui détaille les performances de chaque marqueur pour chaque sous-corpus et pour chaque groupe de marqueurs est donné en Annexe B (p.383).

Ce tableau permet de montrer que les marqueurs sélectionnés sont plus nombreux dans les corpus les plus récents (ce qui n'est pas très étonnant étant donné que la taille des sous-corpus augmente d'une édition à l'autre) mais que les performances atteignent plus de 70% de précision et même 85% dans le sous-corpus TTVS1994. La précision la plus faible est rencontrée dans le TTVS1998 (71%), ce qui encore une fois s'explique par l'hétérogénéité des valeurs que peuvent prendre certains marqueurs très présents dans ce sous-corpus (*plus que,*

meilleur, améliorer) qui ont tendance à renvoyer plus fréquemment à des contextes discursifs spécifiques dans le TTVS1998 que dans les deux autres.

Le Tableau 5.10 ci-dessous détaille pour chacun des sous-corpus les précisions de chaque classe de marqueurs.

	Sous-corpus								
	TTVS1994			TTVS1998			TTVS2002		
	Occur- rences	Perti- nents	Préci- sion	Occur- rences	Perti- nents	Préci- sion	Occur- rences	Perti- nents	Préci- sion
Datation	7	7	100,00	31	29	93,55	53	45	84,91
Nouveauté et disparition de termes/concepts	13	10	76,92	22	19	86,36	34	33	97,06
Nouveaux besoins potentiels	51	43	84,31	124	76	61,29	144	111	77,08

Tableau 5.10 – Précision des marqueurs en fonction de leur classe dans chaque sous-corpus

Dans les trois sous-corpus, les marqueurs « de datation » sont parmi les plus précis, suivis par les marqueurs de « nouveauté et disparition ». Néanmoins, il apparaît également nettement que la performance des marqueurs choisis dépend du sous-corpus, bien que ces variations restent difficiles à expliquer précisément. Ce type d'indice linguistique pour traiter l'évolution doit donc être considéré avec précaution et devra être mis en œuvre dans d'autres types de corpus (et avec d'autres marqueurs également) afin d'étudier au mieux les possibilités qu'il offre. Néanmoins, les performances globales présentées ici restent tout à fait satisfaisantes pour la démarche d'exploration qui est la nôtre et offrent une base solide pour observer qualitativement les informations ainsi extraites.

5.3.2 Observations de contextes

Le premier point de vue possible est l'analyse de chacun des contextes extraits pour retracer leur évolution. Pour ce faire, deux perspectives sont possibles et complémentaires : la synchronie dynamique et la diachronie.

5.3.2.1 Point de vue synchronique dynamique

Une première manière d’aborder l’analyse des contextes extraits est de les observer en synchronie. Plus exactement, le point de vue adopté est un point de vue synchronique dynamique (cf. §3.1.1.1) puisqu’il s’agit de repérer des éléments d’évolution dans un état de connaissances donné dans un corpus synchronique. Dans notre cas, ceci correspond à l’observation des contextes dans chacun des sous-corpus isolément.

Comme les neuf classes de marqueurs définis l’indiquent, les informations observables en contextes peuvent être de différentes natures, mais certaines présentent un intérêt particulier pour apprendre à connaître l’histoire, les étapes ou la chronologie du domaine dont on analyse l’évolution. C’est le cas notamment des datations. En effet, les contextes qui contiennent des marqueurs de datation permettent assez rapidement au linguiste de placer les premiers repères de l’évolution qu’il tente de saisir. Par exemple, dans le TTVS2002 des contextes tels que⁸⁷

(liii) *Depuis quelques années, la technique d’usinage ionique sans contact avec le **miroir** est opérationnelle.*

(liv) *Un produit nouveau est apparu depuis quelques années sur le marché, il s’agit de **multi barrettes**, c’est-à-dire que sur la même **puce détectrice** sont implantées plusieurs **barrettes**.*

permettent de retenir que les multi-barrettes ou la technique d’usinage ionique pour les miroirs sont des concepts récents en 2002, au moment de l’édition du cours. Plus précisément, à l’aide des marqueurs de nouveauté, le premier extrait souligne des progrès récents de la technologie des miroirs, alors que le second indique de nouvelles technologies pour les puces détectrices. Ces extraits illustrent la complémentarité qui existe entre les différentes classes de marqueurs

La datation des informations n’est cependant pas toujours liée à la date de rédaction du corpus, mais plutôt à une chronologie donnée dans les contextes. Par exemple, les extraits

⁸⁷ À partir de cette section, dans les exemples présentés les marqueurs seront soulignés et les termes du domaine présents dans le contexte seront mis en gras.

du TTVS 2002 ci-dessous offrent des éléments intéressants pour retracer l'histoire des lasers dans le spatial :

- (lv) *L'émission dans **un barreau de verre dopé terre rare** a été démontrée dès les années 60, soit au tout début de l'histoire des **lasers**.*
- (lvi) *D'énormes progrès ont été faits dans le domaine des **lasers** depuis le démarrage de ce projet (années 80)*
- (lvii) *Les progrès de la technologie **fibre optique** ont ensuite permis, dans les années 80, d'utiliser des **fibres monomodes** améliorant le **guidage**, donc l'efficacité du **laser**.*

Ces contextes datent l'apparition des lasers dans les années 60 et permettent de dater le début de l'utilisation des fibres monomodes dans les années 1980. Cette période est ainsi marquée comme centrale dans l'histoire et les progrès des technologies laser.

Ceci constitue l'un des intérêts majeurs des contextes qui, associés à une interprétation humaine, facilitent la déduction et la reconstruction d'éléments sur l'évolution du domaine, même si l'analyste n'est lui-même pas expert du domaine qu'il observe. Ceci fait de cette approche une démarche riche pour entamer le repérage de l'évolution d'un domaine pour lequel aucun *a priori* n'est posé.

Un autre exemple est illustré par ces extraits (toujours du TTVS2002) :

- (lviii) *Le premier projet européen d'utilisation de **diodes lasers** sur un satellite est le projet **Silex** (Semi-conductor laser Intersatellite Link EXperiment). Ce projet a pour objet la démonstration d'une **liaison optique inter-satellite** (LEO-GEO) à 0,8 μm . Le premier terminal est monté sur Spot 4 (lancé en 2000), le second sur ARTEMIS (lancé en 2001).*
- (lix) *Les **télécommunications intersatellites** : les résultats récents de la liaison Silex*

L'intérêt de ces contextes est qu'ils renvoient à un projet spécifique, le projet Silex. Ils indiquent que le projet a été mis en application autour de 2000/2001, et qu'il exploite les télécommunications intersatellites. Le champ des télécommunications intersatellites apparaît en évolution, en lien avec les technologies lasers. En effet, il se démarque de projets

plus anciens par l'utilisation de diodes lasers sur les satellites pour la télécommunication. Ce projet représente donc un tournant dans l'évolution des lasers au début des années 2000.

Un second type d'information intéressant à observer en synchronie dynamique sont les remplacements et successions, dans la mesure où ils permettent de suivre les différentes générations d'une même technologie ou d'un même concept, mais dans un corpus synchronique. Ainsi par exemple (TTVS1998) :

- (lx) *Le **pompage optique** se faisait dans un premier temps par **lampe flash**, transversalement, sur des **fibres multimodes**. Les progrès de la technologie **fibre optique** ont ensuite permis, dans les années 80, d'utiliser des **fibres monomodes** améliorant le **guidage**, donc l'efficacité du **laser**. La **fibre** peut alors être utilisée à la fois pour guider l'onde à **amplifier** mais aussi l'onde de **pompage (pompage longitudinal)**. Ceci nécessite toutefois une source puissante et cohérente, avec des qualités de **faisceau** suffisantes. On remplace alors les **lampes** par des **diodes lasers**.*

Dans ce contexte on observe que dans les années 1980 le pompage optique se faisait vraisemblablement à l'aide de fibres monomodes et lampes flash. Les progrès des fibres multimodes ont entraîné le remplacement des fibres monomodes par les multimodes et celui des lampes flash par les diodes lasers.

Ces quelques exemples montrent la richesse des informations que l'on peut mettre au jour par l'observation des contextes en synchronie. Cependant, bien que le dynamisme des connaissances transparaisse en synchronie, il reste souvent intéressant de compléter l'observation par une comparaison diachronique des contextes pour retracer l'évolution des connaissances à l'intérieur de l'intervalle temporel du corpus.

5.3.2.2 Point de vue diachronique

Le point de vue diachronique suppose que l'on dispose d'un corpus comparable, comme c'est le cas pour les corpus constitués dans notre recherche. Cette perspective met en jeu « l'intertextualité » de ce corpus (§3.2.2.1) puisque l'idée sous-jacente dans ce type de comparaison est que les contextes riches en connaissances évolutives sont susceptibles

d'évoluer d'un sous-corpus à l'autre pour marquer l'évolution de concepts. Cette hypothèse est d'autant plus intéressante dans un corpus comme le TTVS qui est composé de rééditions, et donc de mises à jour, d'un même cours.

La comparaison de contextes permet d'accéder à des évolutions suivies telles que celles de la détection infrarouge et du refroidissement. En effet, en 1994, le contexte suivant indique :

- (lxi) *En outre, en utilisant la propriété de couplage radiatif d'une plaque échangeant de l'énergie avec l'espace froid, on réalise des **radiateurs passifs**, capables de **refroidir détecteurs infrarouges**, filtres et optiques associées, à des températures bien stabilisées (quelques degrés) autour de 80 à 100° K. La puissance thermique que peuvent dissiper ces instruments est en constante augmentation, grâce aux récents progrès de **machines cryogéniques actives** (typiquement 500 mW à 1 W autour de 80° K) ; ces machines permettront de **refroidir** des grands plans focaux, constitués par un grand nombre de **détecteurs infrarouges (barrettes ou mosaïques)** et ouvrent ainsi la voie à la haute résolution spectrale dans le domaine des fenêtres atmosphériques thermiques (3 à 5 μm et 8 à 12 μm essentiellement) (TTVS1994)*

Alors qu'en 2002, la même thématique est reprise dans des contextes d'évolution quelque peu différents :

- (lxii) *L'**infrarouge** est toujours lié à la **capacité de refroidir** : optiques et **détecteurs**. Si on arrive à s'affranchir de cette contrainte, on peut simplifier considérablement l'instrument. Or, récemment, de grands progrès ont été accomplis dans le domaine des **mosaïques de détecteurs thermiques non refroidis** (matrices de **microbolomètres**, par exemple). Ces **détecteurs thermiques** (le principe est la mesure d'un échauffement produit par le rayonnement optique) ont une sensibilité spectrale très large (le pouvoir d'absorption de la lumière n'est pas lié à la longueur d'onde) et peuvent être utilisés pour des missions très diverses. Simplicité d'emploi et économie d'un système de refroidissement sont les deux avantages principaux de ces nouveaux instruments infrarouges. Par contre, leur sensibilité est plus faible que celle des **instruments refroidis** utilisant des **détecteurs quantiques**. (TTVS2002)*

Ces contextes indiquent l'importance de la question du refroidissement pour la détection infrarouge et ce, dès 1994. Cependant, si les progrès envisagés en 1994 impliquent l'utilisation de machines cryogéniques, on voit dans l'édition 2002 que la solution finalement envisagée à cette époque est toute autre : plutôt que d'envisager des refroidissements très puissants à l'aide des machines cryogéniques actives, en 2002, la nouvelle solution est d'avoir recours à des systèmes dits « non-refroidis ».

Un second exemple de suivi dans le temps est celui de la technologie BiCMOS :

(lxiii) *En ce qui concerne la **technologie bi-CMOS**, ce n'est qu'en 1993 que nous venons d'avoir accès aux logiciels [...] de conceptions de circuits. (TTVS1994)*

(lxiv) *Aujourd'hui, [...], ce sont les technologies **CMOS** qui font l'objet d'améliorations incessantes. [...] On constate que les technologies **BiCMOS** ne sont qu'une amélioration des technologies **CMOS** existantes. (TTVS2002)*

Ce qui est intéressant dans ce cas est que ces contextes permettent de dater l'apparition des BiCMOS au Cnes en 1993, date où l'accès aux logiciels de conception a été permis, mais ils permettent également de proposer un premier retour sur ces technologies qui, bien que potentiellement intéressantes pour le spatial, ne sont présentées en 2002 que comme des « améliorations » des CMOS (et non pas comme les successeurs des CMOS, comme on aurait pu l'envisager à la lecture de ce dernier contexte).

L'analyse de chacun des contextes fournit donc des informations intéressantes et centrales sur l'évolution des connaissances, et permet entre autres de poser des « jalons » dans la chronologie d'un domaine pour retracer son évolution. Ces contextes mettent en jeu des informations variées et très riches qu'il est possible cependant de compléter par un second type d'analyse : l'observation des termes qui ont tendance à se regrouper dans ces contextes.

5.3.3 Observations de termes récurrents en contextes

Comme nous l'avons défini au début de ce chapitre, les contextes riches en connaissances évolutives contiennent des marqueurs d'évolution, ainsi que des termes du domaine. Il est donc possible de recenser les termes qui tendent à apparaître fréquemment dans ces contextes. Ces termes peuvent être triés de deux manières :

- un tri en fonction de la valeur d'évolution exprimée dans le contexte où ils apparaissent,
- un tri en fonction des sous-corpus dans lesquels ils apparaissent dans ces contextes.

Le premier tri, bien que prometteur, s'avère cependant très délicat à mener. En effet, dans un même contexte d'évolution peuvent coexister différentes valeurs d'évolution et il n'est pas toujours facile d'interpréter clairement une seule évolution pour un terme donné. Par exemple, dans le contexte :

- (lxv) *A la vue des résultats très encourageants obtenus, trois programmes aujourd'hui envisagent la conception des chaînes vidéo en **ASIC**. Il s'agit de la nouvelle génération des **satellites d'observation de la terre** (Spot5 et HELIOS2) et de l'instrument **IASI, interféromètre** chargé de l'analyse de l'**atmosphère** à des fins météorologiques qui a la particularité d'être demandeur d'électronique (conçue en ASIC) à très faible consommation devant fonctionner à 100 K.*

certain satellites sont marqués comme nouveaux (les satellites Spot5 et HELIOS2) mais il est difficile d'interpréter avec précision l'évolution liée au terme/concept ASIC : s'agit-il d'un concept récent ? Les ASIC sont-ils connus depuis longtemps, mais disponibles seulement récemment dans le spatial ? Ont-ils connu des améliorations qui entraînent ces résultats encourageants ? Si oui, lesquelles ? De la même manière, *atmosphère* est un terme du domaine extrait par Syntex et présent dans ce contexte auquel on ne peut lier aucune évolution.

Pour cette raison, nous adoptons la seconde possibilité de tri et observons les termes récurrents dans les contextes riches en connaissances évolutives dans chacun des trois sous-corpus du TTVS.

Ce type d'observation permet de repérer par exemple qu'en 1998, des termes liés au projet Silex sont récurrents dans les contextes d'évolution : *constellations de satellites, constellations, liaison intersatellite, polynômes*. Ces termes permettent de mettre en avant l'importance de ce projet dans l'évolution de l'optique au milieu des années 1990.

Mais l'un des aspects les plus marqués et intéressant dans ce type de démarche est qu'il est possible de retrouver des thématiques déjà identifiées à l'aide de l'indice de fréquence

(§4.3.2, p.132). Ainsi, parmi les termes fréquents qui apparaissent dans des contextes évolutifs uniquement en 1994, on trouve les termes *machine cryogéniques* et *radiateur* qui appartiennent à la thématique du refroidissement, dont nous avons souligné l'évolution *supra*. On retrouve encore les termes disparus *synthèse d'ouverture optique* et *ouverture optique* qui renvoient à un concept jugé prometteur au début des années 1990, mais qui est toujours en développement pour le spatial.

En 2002, on retrouve encore une fois une thématique déjà mentionnée à plusieurs reprises, notamment dans le cadre des regroupements de formes chrono-homogènes : la thématique des miroirs et la question des parasites. Ainsi, les termes *arêtes*, *couches minces optiques*, *couche mince*, *miroir asphérique*, *miroir plan*, *miroir primaire* ou encore *flux parasite*, *image parasite* sont récurrents dans les contextes évolutifs du TTVS 2002.

Cependant, ces termes n'ont pas tous le même statut. En effet, dans cet exemple, les termes *miroir asphérique*, *miroir plan*, *miroir primaire* ne sont jamais les « sujets » de l'évolution, mais plus des termes/concepts anciens auxquels on fait référence dans ces contextes parce qu'ils sont liés à une évolution. Par exemple, dans le TTVS 2002 :

(lxvi) *Depuis quelques années, la technique d'usinage ionique sans contact avec le miroir est opérationnelle. Un faisceau d'ions vient arracher la matière. Malheureusement, cette technique ne peut être utilisée qu'en phase finale car un faisceau de trop grande énergie provoque un échauffement important du miroir ; ainsi le taux d'arrachement de matière est faible. Toutefois, elle offre l'avantage d'être très déterministe et laisse entrevoir la possibilité de corriger les défauts de quilting générés par les phases précédentes. S'affranchir des contraintes de quilting dans le dimensionnement est la solution pour un gain de masse précieux pour les grands miroirs envisagés dans le futur.*

Dans ce contexte, on compte 3 occurrences du terme *miroir*. Cependant, le terme/concept *miroir* n'y est pas présenté comme cible de l'évolution. Ce contexte présente plutôt des évolutions liées à l'usinage ionique (donc au polissage) des miroirs. De fait, ce n'est pas directement le concept *miroir* dont on décrit l'évolution, mais plutôt celle des techniques qui lui sont associées. Cependant, la présence de ce terme n'est pas anodine dans ces contextes puisqu'il s'agit d'un terme/concept ancien que l'on rappelle pour aborder la question des progrès dans lesquels il est impliqué. Autrement dit, des termes/concepts connus ou anciens

qui gravitent dans certains contextes d'évolution pourraient être considérés comme des marqueurs « indirects » de l'évolution d'autres termes/concepts. Nous retrouvons donc ici une remarque amorcée au chapitre précédent.

5.4 Bilan

Dans ce chapitre, nous avons mis en avant le potentiel de techniques d'extraction d'information pour la détection de contextes riches en connaissances évolutives. Malgré certaines difficultés à caractériser et définir les marqueurs, ce type de démarche présente des intérêts méthodologiques certains pour observer l'évolution des connaissances à partir de textes.

Le premier est l'adaptabilité de cette approche à différents types d'informations recherchés et la diversité des informations que l'on cherche à extraire (nouveau obsolescence, datation, amélioration, etc.).

De plus, les contextes permettent de repérer des évolutions sur au moins deux types différents d'éléments : les concepts du domaine et leurs propriétés. Par exemple dans un contexte tel que :

(lxvii) Ces nouvelles technologies n'ont pas encore réellement fait leur entrée dans le spatial, car elles sont plus récentes que celle des **semi-conducteurs**. Elles ont donc encore à prouver leurs potentialités dans ce domaine. Mais il est certain que leur évolution est à suivre de près pour les futures applications spatiales, où faibles **encombrement** et **masse**, faible **consommation électrique** et bas coût seront des caractéristiques déterminantes. (TTVS1998)

on extrait à la fois des informations sur l'évolution de concepts (les nouvelles technologies citées en début de paragraphe et les semi-conducteurs), ainsi que sur les propriétés recherchées dans les futures applications spatiales : faible encombrement, faible masse, faible consommation, bas coût. Cet élément est important dans la mesure où ces deux types d'informations sont susceptibles d'être centraux dans certaines applications pour lesquelles on met en place un repérage de l'évolution. Par exemple, dans le cas de la mise à jour de ressources termino-ontologiques, il peut s'avérer important de repérer à la fois les

évolutions des concepts, mais également de leurs propriétés. Néanmoins, il reste difficile de nuancer ces deux aspects à ce stade dans la mesure où ils sont marqués de manière semblable dans les contextes.

Le deuxième intérêt méthodologique de cette approche est que l'analyse des contextes s'avère une étape riche pour commencer l'exploration de l'évolution d'un domaine pour lequel on ne dispose que de peu de connaissances. En effet, l'observation de contextes est relativement facile à mettre en œuvre (dépendamment de la finesse de l'extraction souhaitée), et permet à l'analyste de s'immerger directement et globalement dans les textes et dans le domaine. Il lui est alors facile de dégager des éléments dans la chronologie du domaine par déduction, même s'il n'est pas expert de ce domaine et s'il ne dispose que de peu d'informations *a priori*. L'analyse manuelle de ce type de contextes, que ce soit en synchronie (dynamique) ou en diachronie, lui permet également de travailler à partir de pans entiers de textes qui traitent explicitement de l'évolution du domaine observé et de s'immerger ainsi directement dans les corpus et les connaissances.

Le recours aux contextes a permis également de confirmer l'hétérogénéité des informations relatives à l'évolution et repérables en diachronie courte. À l'instar des premiers éléments mis en avant dans le chapitre précédent, il est clair que l'évolution des connaissances fait appel à un panel varié de types d'évolution. Néanmoins, ce deuxième indice permet de postuler une tendance intéressante pour l'observation : les indices peuvent se combiner pour indiquer une même évolution. Ainsi, l'évolution de la thématique du refroidissement, déjà repérée par l'indice de fréquence, est également mise en avant grâce aux contextes riches en connaissances évolutives.

Il reste cependant beaucoup de perspectives pour affiner cet indice, la première étant la définition d'autres marqueurs. Dans cette optique, il serait particulièrement intéressant de prendre en compte la dimension discursive et la structure des textes (Laignelet, 2009 ; Péry-Woodley, 2000 ; Toussaint, 2004). En effet, il apparaît clairement dans nos corpus que certaines sections des textes sont propices à l'extraction de contextes, en particulier les

introductions, les conclusions et les sections dont le titre contient lui-même un marqueur d'évolution. De plus, sous réserve de les caractériser plus précisément, on pourrait envisager de mettre au jour des configurations de marqueurs en fonction des évolutions à repérer (Ho-Dac & Péry-Woodley, 2008 ; Laignelet, 2009 ; Péry-Woodley, 2000).

Ces descriptions offrent donc de nombreuses pistes pour l'exploration de l'évolution. Dans la suite de cette partie, nous focaliserons notre attention sur la définition de deux autres indices qui permettront entre autres de développer la question de l'hétérogénéité des interprétations possibles en diachronie courte, ainsi que la tendance des indices à se regrouper pour pointer vers une même évolution. Pour cela, le prochain indice observé est celui des variantes de termes.

Chapitre 6 Variantes terminologiques

SOMMAIRE DU CHAPITRE

6.1	VARIANTES TERMINOLOGIQUES ET DIACHRONIE.....	190
6.1.1	<i>Définition et richesse des perspectives diachroniques.....</i>	<i>190</i>
6.1.2	<i>Entre langue et discours</i>	<i>193</i>
6.2	EXTRACTION ET SÉLECTION DES VARIANTES TERMINOLOGIQUES	196
6.2.1	<i>Typologie définie pour cette étude</i>	<i>196</i>
6.2.2	<i>Sélection des variantes.....</i>	<i>199</i>
6.3	OBSERVATION EN CORPUS	206
6.3.1	<i>Évaluation de la pertinence des variantes terminologiques en diachronie.....</i>	<i>206</i>
6.3.2	<i>Apparition simultanée de variantes.....</i>	<i>207</i>
6.3.3	<i>Apparition/disparition de certaines variantes</i>	<i>209</i>
6.3.4	<i>Variante lexicales et variantes discursives : recours aux contextes pour l'analyse.....</i>	<i>214</i>
6.4	BILAN.....	219

Dans ce chapitre, nous présentons un troisième indice pour observer l'évolution des connaissances : les variantes terminologiques. Les variantes sont un phénomène central traité en terminologie et en TAL depuis une quinzaine d'années et offrent de riches perspectives d'un point de vue diachronique. Il s'agit cependant d'un phénomène linguistique difficile à traiter dont l'une des spécificités est d'intervenir soit au niveau lexical soit au niveau discursif, niveaux complémentaires pour l'analyse diachronique.

Dans la première section, nous présentons une description des points de vue diachroniques sur les variantes et montrons la particularité des variantes entre lexique et discours. Dans la deuxième section nous décrivons la typologie et l'extraction des variantes pour cette analyse. Enfin, nous terminons par une description des phénomènes d'évolution observables à l'aide des variantes et proposons des pistes de réflexion pour prendre en compte les deux niveaux, lexical et discursif.

6.1 Variantes terminologiques et diachronie

6.1.1 Définition et richesse des perspectives diachroniques

La variation terminologique est définie par Freixa (2006 : 51) comme le fait qu'un concept puisse être dénommé par différents termes de manière stable et consensuelle parmi les experts d'un domaine :

« Denominative variation can be defined as the phenomenon in which one and the same concept has different denominations; this is not just any formal variation (variation between a term and a periphrasis or a definition, for example), but is restricted to variation among different denominations, i.e., lexicalised forms, with a minimum of stability and consensus among users of units in a specialised domain. »

En diachronie, plusieurs perspectives mettent à contribution la question des variantes pour l'analyse. Parmi les premiers intérêts des variantes, on trouve le fait qu'elles permettent de postuler une relation dynamique entre le terme (dénomination) et le concept. C'est l'optique avancée par exemple par Faulstich (1998/1999) qui propose une analyse fonctionnelle de plusieurs termes/concepts en portugais pour souligner cette dynamique, d'abord en synchronie mais surtout en diachronie. L'objectif de l'auteur est de montrer à travers les variantes que les terminologies ont une évolution et une histoire propre.

La dynamique entre le terme et le concept a inspiré également certains travaux de TAL, dont ceux de Tartier (2003 ; 2004 ; 2006 ; 2007) qui pose l'hypothèse d'un parallèle entre évolution de la dénomination et évolution du concept. L'idée générale de ses travaux est que, d'un point de vue formel, plus une variante est éloignée d'une autre, plus le lien conceptuel qui les relie a changé. L'auteur tente ainsi de caractériser automatiquement ce parallélisme pour repérer des évolutions dans un domaine donné (cf. §1.2.3.3, p.35) et traduit ainsi de manière automatique le parallèle minimal entre évolution de la dénomination (interne) et évolution du concept (externe).

Un autre point de vue est celui de la veille scientifique et technique, illustré notamment par les travaux d'Ibekwe-Sanjuan et Sanjuan (Ibekwe-Sanjuan, 1998 ; Ibekwe-Sanjuan &

Sanjuan, 2002). Les variantes terminologiques sont considérées dans ce cadre comme un moyen de repérer les changements de thématiques dans les textes. Dans la mesure où leur recherche s'inscrit dans une optique de veille, les deux auteurs accordent une importance toute particulière aux variantes dites syntaxiques, telles que les expansions (*root of bragg* → *root exudate of soyabean cultivar bragg*). Ce type de variantes permet en effet d'extraire des relations de type générique/spécifique qui, lorsqu'elles apparaissent en corpus, leur permettent alors d'identifier non seulement des thématiques émergentes mais aussi leur organisation spécifique.

Mais de manière plus générale, la question des variantes est associée en diachronie à deux types de phénomènes : le foisonnement terminologique et l'implantation.

Le premier phénomène, le foisonnement terminologique est « caractéristique des domaines novateurs » (Boutayeb, 1996) et intervient lors de l'apparition de nouveaux concepts (Guilbert, 1965). Déjà Guilbert (Guilbert, 1965 ; 1973a ; 1975) avait proposé la notion de « vocabulaire en gestation » pour décrire l'étape d'apparition d'une notion dans un domaine, notion reprise et illustrée récemment par Dury (1999 ; 2008a ; Dury & Lervad, 2007) dans les domaines de l'écologie et du pétrole. Cette étape est généralement marquée au niveau de la terminologie par un foisonnement de variantes concurrentes pour dénommer la notion en train d'apparaître. C'est cette profusion de variantes que l'on appelle « foisonnement terminologique ». De ce point de vue, l'apparition simultanée de plusieurs variantes concurrentes pour un même concept à un moment donné dans un corpus diachronique peut permettre d'émettre l'hypothèse que ce foisonnement implique l'apparition d'un nouveau concept dans le domaine.

À ce phénomène peut être ajouté un second, celui de l'implantation. Foisonnement et implantation peuvent être complémentaires dans la mesure où, comme le souligne Guilbert (1975 : 11-12), la coexistence de variantes concurrentes est généralement temporaire :

« S'il arrive que plusieurs termes soient employés pour désigner une même chose la distinction ne porte que sur le signifiant, les éléments de signification étant exactement calqués sur la chose. La synonymie porte sur la dénotation ou dénomination. Elle a d'ailleurs la plupart du temps un caractère très

provisoire ; elle correspond à la période de recherche et d'élaboration d'un concept ou d'une invention. »

Plus précisément, à la suite de l'apparition simultanée de plusieurs variantes, parmi les variantes concurrentes, l'une d'entre elles peut terminer son intégration dans la terminologie et s'implanter au détriment des autres, qui finiront probablement par disparaître. Autrement dit, lorsque l'on observe un foisonnement de variantes terminologiques pour un même concept, avec le temps ce foisonnement diminue pour laisser place à de moins en moins de variantes dont on dit alors qu'elles sont implantées.

Le phénomène d'implantation est central dans le cadre de la normalisation en terminologie, comme décrit au Chapitre 1 : l'étude de l'implantation est au cœur d'études terminologiques ou terminométriques telles que (Depecker, *et al.*, 1997 ; Quirion, 2003a ; 2003b). En effet, comme le souligne Quirion (2004 : 193), les changements terminologiques peuvent être de deux ordres : planifiés, dans le cas de politiques d'aménagement linguistique, ou spontanés, dans le cas de l'évolution libre de la langue. La planification d'un changement en terminologie est liée à la mise en place d'une politique linguistique d'aménagement qui vise à éliminer des variantes « non désirables » (par exemple les variantes anglaises au Québec où la politique linguistique vise la valorisation du français) ou un trop grand nombre de variantes concurrentes. Une fois les propositions d'aménagement appliquées, il importe d'évaluer dans quelles mesures ces termes proposés sont finalement implantés dans la terminologie du domaine visé et si la politique a porté ses fruits. Dans cet objectif, plusieurs protocoles d'évaluation d'implantation sont proposés, dont le logiciel Barçah développé par Quirion (2003a). Dans le cas où le changement est non planifié, le même type d'évaluation peut être mis en place pour observer laquelle ou lesquelles des variantes fini(ssen)t par s'implanter dans la terminologie du domaine, bien qu'à notre connaissance aucun travail de ce type n'existe.

Notons pour finir que foisonnement et implantation sont complémentaires, à la fois au niveau linguistique et au niveau extralinguistique. En effet, si d'un point de vue linguistique (interne) le foisonnement indique la coexistence de plusieurs dénominations concurrentes dont on cherche à savoir laquelle va s'implanter, d'un point de vue externe le foisonnement peut indiquer la nouveauté – ou à défaut la récence – d'un concept ; l'implantation quant à

elle peut indiquer la familiarisation, la stabilisation ou la maturité de ce même concept dans le domaine. Nous y reviendrons dans la section 6.3 de ce chapitre.

6.1.2 Entre langue et discours

Bien que les travaux mentionnés ci-dessus n'exploitent pas cette distinction, la richesse des variantes repose en partie sur le fait que ces dernières mettent en jeu une distinction centrale entre niveau lexical et niveau discursif. En effet, malgré le fait que les travaux mentionnés ci-dessus analysent les variantes comme des unités terminologiques nouvelles ou émergentes, les variantes sont aussi un phénomène discursif où l'on ne met plus en jeu des unités terminologiques concurrentes, mais une unité terminologique et sa variante utilisée en tant que reprise anaphorique au fil du discours. Cette distinction est illustrée en particulier dans le cas de variantes obtenues par *réduction*, c'est-à-dire par la suppression d'un des composants de la variante, généralement un de ses modificateurs.

La réduction de termes peut être due à deux phénomènes : une réduction qui relève d'une reprise anaphorique, une réduction à caractère lexical (Collet, 2002 pour une description détaillée ou Alber-Dewolf, 1982 ; Collet, 1997 ; Jacques, 2003 ; Kocourek, 1982 ; Quemada, 1978)⁸⁸.

Dans le premier cas, la réduction est une reprise anaphorique d'un terme complexe dans un discours et participe à la cohésion et cohérence du texte. Dans le second cas, la variante est une unité lexicale autonome reconnue, généralement équivalent sémantique du terme complexe dont elle est la réduction. Cette différence est bien établie en linguistique, et plusieurs indices ont été décrits pour les distinguer, indices qui mettent notamment en avant le statut d'indépendance contextuelle des réductions à caractère lexical par rapport aux reprises anaphoriques, ainsi que leur indépendance par rapport aux termes complexes dont elles sont issues (Alber-Dewolf, 1982 ; Collet, 2002 ; Jacques, 2003).

⁸⁸ Dans la littérature, on trouve notamment les termes de « réduction anaphorique (RA) et « réduction à caractère lexical (RL) » chez Collet (2002) mais d'autres dénominations existent pour ce phénomène dont *ellipse contextuelle* et *néonymique* chez Alber-Dewolf ou *ellipse contextuelle* et *lexicale* chez Kocourek. Pour notre part, nous parlerons de variante discursive et de variante lexicale.

Du point de vue diachronique, en particulier en néologie, la description des réductions à caractère lexical est privilégiée :

« Par opposition à l'ellipse contextuelle, l'ellipse néonymique se caractérise par son autonomie, c'est-à-dire son indépendance par rapport au contexte sur le plan sémantique (ex. Fr voiture automobile → automobile [...]). Et c'est l'ellipse néonymique seule qui intéresse la recherche en néonymie. L'ellipse néonymique est un procédé couramment utilisé en Lsp pour réduire des UTC longues et encombrantes. » (Alber-Dewolf, 1982 : 37)

Mais dans les faits, ces deux types de réductions peuvent être considérés dans une certaine continuité. C'est la piste proposée notamment par Jacques (*ibidem* : 248), à la suite de Bonhomme (1987) :

« Leurs [les termes réduits] caractéristiques actuelles correspondent à **une étape de la construction des unités lexicales du domaine. La lexicalisation observée ici s'inscrit dans un processus qui est analysé par M. Bonhomme (1987 : 268) comme « néologie par le discours »**. A la différence de la « néologie de langue » qui crée de nouveaux signifiants en employant les ressorts de la dérivation affixale, la néologie de discours « agit à partir de l'actualisation des phrases-types à la superficie des énoncés » (*ibidem*). **Elle prend comme point de départ les formes réalisées en discours et les intègre au lexique en oubliant les mécanismes discursifs qui les font naître.**⁸⁹ »

La successivité entre reprise anaphorique et réduction à caractère lexical n'est pas facile à observer et relève de ce que l'on pourrait appeler un « cycle d'évolution » qui ne peut vraisemblablement être capté dans toute sa finesse qu'en diachronie longue, comme nous le verrons plus loin.

Dans les études existantes, le point de vue adopté sur le phénomène de variation par réduction est souvent un point de vue socio-historique, qui vise l'explication de ce phénomène sur la base d'éléments extralinguistiques (telles que les évolutions de la société). Ce point de vue peut être illustré à travers deux exemples d'analyses, l'un de Guilbert et l'autre de Portelance, souvent cités dans la littérature sur la question.

⁸⁹ C'est nous qui soulignons.

Guilbert (1975 : 273-274) explique en effet le changement du terme *voiture automobile* en sa réduction *automobile* comme suit :

« Le type d'abréviation de composé le plus répandu consiste à réduire le composé formé du syntagme N + Adj. Au second terme adjectif : [...], une voiture automobile > une automobile [...] On remarque que le terme maintenu est le déterminant [...] Le pouvoir de désignation se situe dans le déterminant et non dans la base à valeur générale. Cette vocation du déterminant à représenter le syntagme découle du fait de sa fonction et non de sa place. [...] Il peut arriver que l'unité syntagmatique de forme nom + adj se réduise à la base : voiture automobile > voiture. L'abréviation n'est plus, alors, d'essence seulement syntagmatique ; **son apparition implique que l'évolution intervenue sur le plan du référent interdit toute ambiguïté, le mode de locomotion le plus répandu étant celui de l'automobile**⁹⁰. »

Autre exemple, Portelance (1996 : 104-105) offre quant à elle une description de la réduction du terme *cigarette* sur la base d'arguments d'ordre sociohistorique :

« Lors de l'apparition de la cigarette sur le marché, l'objet ne comporte pas de filtre, le référent ne s'oppose alors qu'à des objets tels cigare, pipe etc. Par la suite, on introduit une nouvelle cigarette ayant comme particularité un bout filtre ; cigarette filtre est une réduction de cigarette à bout filtre. Dans une situation de concurrence, *cigarette* devient l'hyperonyme de *cigarette à bout filtre* et de *cigarette à bout uni*, ce dernier objet est ainsi renommé pour les besoins de sous-catégorisation. Puis, les habitudes de consommation évoluant, le deuxième type de cigarette devient le plus répandu si bien qu'il est désigné maintenant par le générique [...]. »

La distinction entre ces deux niveaux est importante en diachronie dans la mesure où elle permet de chercher à retracer ce passage entre niveau discursif et niveau lexical. Or, comme nous le verrons dans la suite de ce chapitre, cette distinction ne doit pas être limitée à ce phénomène et ces deux niveaux peuvent également être pris en compte indépendamment pour décrire l'évolution des connaissances. En effet, en diachronie courte on n'observe que très peu d'exemples de passage entre niveau discursif et niveau lexical, processus qui demande des intervalles de temps plus longs. Néanmoins, la distinction entre ces deux

⁹⁰ C'est nous qui soulignons.

niveaux peut avoir un impact d'un point de vue diachronique et impliquer l'analyse de différents types de changements.

Mais avant de décrire cet aspect, dans la mesure où les variantes terminologiques peuvent correspondre à différents types et pour parvenir à une description adéquate de leur apport en diachronie, il convient de caractériser les variantes à prendre en compte.

6.2 Extraction et sélection des variantes terminologiques

6.2.1 Typologie définie pour cette étude

Les types de variantes terminologiques étudiés sont nombreux, ce qui est lié en partie à l'engouement pour ces dernières depuis une quinzaine d'années, en particulier dans les recherches en TAL (notamment Daille, 2005 ; Daille, *et al.*, 1996 ; Jacquemin, 1997 ; 2001 ; Zweigenbaum & Grabar, 2000, et bien d'autres). La notion de *variante terminologique* a été mise à contribution dans de nombreuses applications telles que la recherche d'information, la traduction automatique, la veille scientifique et technologique, etc. (Daille, 2005), multipliant les points de vue possibles sur ce phénomène. Dans cette perspective, Daille *et al.* (1996 : 2) définissent par exemple les variantes terminologiques comme « [...] utterance[s] which [are] semantically and conceptually related to an original term ».

Cette définition n'est pas très éloignée de celle de Freixa donnée *supra* (p.190), si ce n'est dans la description du lien conceptuel entre variantes. En effet, Freixa parle avant tout d'équivalence conceptuelle, Daille propose plutôt de considérer tout lien conceptuel « quel qu'il soit ». Plus précisément, celui-ci peut être défini en fonction des applications dans lesquelles les variantes sont impliquées (Daille, 2005 : 183) :

« several interpretations could be given [...]: a variant could be a synonym of an authorised term, reflect a semantic distance from the reference term, or refer to another term linked to the authorised term by a conceptual link. [...] the adopted definition of term variation is highly dependent on the foreseen application. »

Ainsi par exemple, dans les travaux de Ibekwe-Sanjuan et Sanjuan décrits *supra*, les variantes considérées par les auteurs ne présentent pas d'équivalence conceptuelle, mais entrent dans une relation de type générique/spécifique, essentielle pour repérer l'apparition de nouvelles thématiques dans une optique de veille scientifique.

Dans une certaine mesure, on peut dire que c'est cette « flexibilité » dans la définition des variantes qui leur confère un grand intérêt pour autant d'applications et de points de vue. Néanmoins, pour la diachronie, nous adopterons la définition de Freixa et considérons que les variantes doivent entretenir entre elles un lien d'équivalence conceptuelle.

Si nous reprenons l'exemple de l'expansion comme mécanisme de variation, il semble clair qu'il n'y a généralement pas d'équivalence conceptuelle entre les variantes. L'ajout d'un élément à un terme donné entraîne généralement une spécification de celui-ci et ne relève donc pas pour nous de la variation terminologique mais des dépendances syntaxiques, abordées dans le chapitre suivant. Par exemple si l'on observe la variante *télescope de Korch* obtenue par expansion à partir du terme *télescope*, il n'y a pas d'équivalence conceptuelle dans la mesure où un télescope de Korch est plus spécifique, il s'agit d'un type de télescope. Pour sélectionner des types de variantes terminologiques en lien d'équivalence conceptuelle exclusivement, nous nous alignons sur la typologie proposée par Freixa (2002 : 298sq.), complète et facile à mettre en œuvre.

Dans cette typologie, l'auteure distingue cinq catégories de variation terminologique :

1. Variation par changements graphiques,
2. Variation par changements morphosyntaxiques,
3. Variation par réduction,
4. Variation par changements lexicaux,
5. Variation par changements complexes.

La première catégorie, « **Variation par changements graphiques** », regroupe⁹¹ :

⁹¹ Le choix du terme qui donne naissance aux variantes, c'est-à-dire la sélection de la variante à partir de laquelle on classe les autres variantes, n'est pas trivial. Nous avons choisi de considérer comme « variante source » : les dénominations qui apparaissent dans les corpus les plus anciens et/ou qui comptent le plus d'occurrences ou, par défaut lorsque les deux premiers critères ne sont pas discriminants, les dénominations les plus longues. La seule exception concerne les variantes obtenues par réduction où l'on choisit comme variante de base la variante la plus longue. Par exemple, dans le corpus TTVS, on recense trois variantes qui apparaissent simultanément

- les changements orthographiques : absence/présence de marques typographiques, alternance majuscules/minuscules, etc. Par exemple⁹² :
 - *rapport signal sur bruit* | *rapport signal-sur-bruit*, *ASIC* | *asic*
- les abréviations (majoritairement les troncations)
 - *amplificateur* | *Ampli*
- et les sigles
 - *balise maîtresse* | *BM*.

Dans la deuxième catégorie, « **Variation par changements morphosyntaxiques** », Freixa distingue deux sous-catégories :

a) les changements qui ne modifient pas la structure du terme :

- les changements de préposition
 - *reçu sur le détecteur* | *reçu par le détecteur*
- les changements flexionnels
 - *coffret de batterie* | *coffret de batteries*

b) les changements qui entraînent une modification de la structure :

- la suppression d'une unité grammaticale telle que le déterminant, la préposition, etc.
 - *coffret de batterie* | *coffret batterie*
-
- les transformations du type [SPrép] → [SAdj]
 - *liaison entre satellites* | *liaison intersatellite*

Dans la troisième catégorie, « **Variation par réduction** », la variante de base est un terme complexe et les variantes naissent de la suppression d'au moins un de ses éléments, à savoir :

- la tête
 - *rapport S/B* | *S/B*

dans le corpus : *TS*, *Tracking Sensor* et *capteur de poursuite*. Dans la mesure où *TS* est le sigle anglais et que sa forme développée est présente, c'est cette dernière que nous choisissons comme terme de base. *TS* est considéré comme « siglaison » et *capteur de poursuite* comme « Traductions ».

⁹² Les exemples de variantes présentés sont tirés de nos deux corpus.

- l'expansion
 - *flux lumineux parasite* | *flux parasite, spécification technique du coffret* | *spécification du coffret*.

La quatrième catégorie, « **Variation par changements lexicaux** », prend en compte la substitution d'un élément de la variante (terme simple ou terme complexe) par un synonyme :

- substitution de la tête par un synonyme (aucun exemple dans nos corpus)
- substitution de l'expansion
 - *coffret de batterie* | *coffret d'alimentation*
- substitution du terme entier
 - *panne* | *anomalie*.

La cinquième catégorie, « **Variation par changements complexes** », regroupe les combinaisons de différents types de changements, tels que :

- réduction + abréviation
 - *émission des blocs de téléchargement* | *émission des TCHs, rapport signal-sur-bruit* | *S/B*
- abréviation + changement graphique
 - *amplificateur opérationnel* | *ampli-op*.

Nous rajoutons à la typologie de Freixa une dernière catégorie, « Traductions », qui regroupe les cas de variantes traduites. Par exemple :

- *Coarse Pointing Mechanism* | *Mécanisme de Pointage Gros*.

6.2.2 Sélection des variantes

La sélection des variantes doit prendre en compte deux éléments. Tout d'abord, le fait qu'aucune des catégories de cette typologie n'est, ni ne peut être, associée *a priori* à des types d'évolution des connaissances en diachronie courte. Ensuite, le fait que pour avoir une pertinence en diachronie, les variantes terminologiques ne peuvent être prises en compte

isolément. Les variantes doivent en effet être associées à l'indice de fréquence pour pouvoir observer leur mouvement dans un corpus diachronique et leur faire « acquérir » ainsi une dimension diachronique. En effet, sans empreinte de fréquence spécifique, les variantes terminologiques observables permettent uniquement de signaler la présence de dénominations équivalentes pour un concept donné. Or, ce que l'on cherche à saisir en diachronie ce sont plutôt des informations sur l'évolution de ces variantes dans le temps, par exemple si elles indiquent une évolution du concept qu'elles dénomment.

Pour ce faire, dans notre recherche, les variantes terminologiques extraites sont associées aux empreintes de fréquence suivantes :

1. l'apparition de l'ensemble des variantes dans les sous-corpus les plus récents uniquement,
2. l'apparition ou disparition d'au moins une des variantes, mais non la totalité d'entre elles, dans au moins l'un des sous-corpus.

Soulignons que les empreintes de fréquence ne se basent pas sur le χ^2 comme décrit au Chapitre 4 (p.111) dans la mesure où les variantes ne comptent souvent que très peu d'occurrences, ce qui ne permet pas d'appliquer le χ^2 . De fait, les deux empreintes choisies sont appliquées simplement sur le nombre d'occurrences recensé dans chaque sous-corpus (Tableau 6.1 et Tableau 6.2).

La première empreinte permet de repérer des groupes de variantes qui apparaissent simultanément, à la manière du foisonnement terminologique décrit *supra* (Tableau 6.1). Néanmoins, comme nous le verrons, le terme « foisonnement » n'est pas pleinement approprié pour qualifier le phénomène d'apparition de variantes en diachronie courte. En effet, lorsque Guilbert (1965) évoque le foisonnement dans sa description de la « gestation » de terminologies, il inscrit ce phénomène dans le champ de la néologie et parle spécifiquement de « foisonnement néologique transitoire » :

« Dans la période de création d'une réalité nouvelle et de formation d'un vocabulaire adéquat, c'est une caractéristique de la situation linguistique qu'un certain foisonnement néologique transitoire se produise pour désigner un même concept. » (*ibidem* : 331, cité par (Dury, 2008a : 14))

Dury (2008a) propose quant à elle de parler plutôt de « synonymie de circonstance » dans la mesure où elle observe que

« ce foisonnement se manifeste par la mise en concurrence provisoire de néologismes **tout autant que de termes existant déjà dans le domaine**⁹³, mais aux contours sémantiques assez flous, qui deviennent alors concurrents, de façon transitoire »

Le foisonnement peut donc être décrit comme un phénomène de concurrence entre variantes lors de la création d'un concept nouveau. Or, comme nous le verrons, en diachronie courte on ne peut observer de phénomène lié à la naissance d'un concept aussi marqué. Pour cette raison, nous gardons le terme générique de « concurrence » pour parler de ce phénomène et réservons « foisonnement » aux cas de néologie propres.

Variantes	Fréquence DORISGen1	Fréquence DORISGen3
<i>visibilité de satellite</i>	0	4
<i>visibilité satellite</i>	0	8

Tableau 6.1 – Exemple de groupe de variantes en concurrence (DORIS)

La seconde empreinte permet de repérer les cas de succession ou de remplacement de variantes existantes pour un concept donné (Tableau 6.2). En d'autres termes, cette empreinte permet de repérer des cas qui se rapportent à l'implantation décrite *supra*.

Variantes	Fréquence TTVS1994	Fréquence TTVS1998	Fréquence TTVS2002
<i>rapport signal sur bruit</i>	2	6	7
<i>rapport signal à bruit</i>	10	9	8
<i>rapport signal-sur-bruit</i>	0	0	1
<i>rapport S/B</i>	2	9	14
<i>rapport signal/bruit</i>	1	2	1
<i>S/B</i>	6	15	9
Total	21	41	40

Tableau 6.2 – Exemple de groupe de variantes dont au moins une apparaît/disparaît (TTVS)

⁹³ C'est nous qui soulignons.

Dans ce cas, il est intéressant d'appliquer le coefficient d'implantation proposé par Quirion (2003a ; 2003b) pour mesurer l'implantation des termes. Ce calcul terminométrique permet de calculer pour chaque variante d'un concept sa proportion d'emploi dans les textes par rapport à l'ensemble des variantes qui existent (Figure 6.1).

$$\frac{\text{Nombre d'occurrences d'une variante donnée}}{\text{Somme de l'ensemble des occurrences de toutes les variantes de la famille}} \times 100$$

Figure 6.1 – Coefficient d'implantation⁹⁴ (Quirion, 2003a ; 2003b)

Appliqué à l'exemple du Tableau 6.2, ce calcul permet d'obtenir les résultats suivants (Tableau 6.3) :

Variantes	Coefficient d'implantation		
	TTVS1994	TTVS1998	TTVS2002
<i>rapport signal sur bruit</i>	9,52	14,63	17,50
<i>rapport signal à bruit</i>	47,62	21,95	20
<i>rapport signal-sur-bruit</i>	0	0	2,50
<i>rapport S/B</i>	9,52	21,95	35
<i>rapport signal/bruit</i>	4,76	4,88	2,50
<i>S/B</i>	28,57	36,59	22,50
Total	100	100	100

Tableau 6.3 – Exemple de calcul du coefficient d'implantation (TTVS)

Dans cet exemple, le calcul du coefficient permet de montrer que pour le même concept, la variante la plus fréquente dans le TTVS en 1994 est *rapport signal à bruit* alors qu'à partir de 1998, les variantes privilégiées sont les formes les plus courtes : *rapport S/B* et *S/B*. À partir de ces résultats, on peut chercher à savoir si ce concept s'est lui aussi stabilisé dans le domaine où si l'évolution observée ne concerne que sa dénomination.

À partir de la typologie des variantes terminologiques proposée et des empreintes de fréquence choisies, les variantes sont extraites de nos deux corpus. Les recherches en TAL se sont beaucoup penchées sur la question des variantes et il existe aujourd'hui plusieurs outils automatiques de repérage pour le français tels que FASTR (Jacquemin, 2001) ou ACABIT (Daille, 2003). Cependant, ces deux outils ne se sont pas révélés exploitables pour notre

⁹⁴ Dans son calcul original, Quirion propose un coefficient en 0 et 1. Nous le convertissons en pourcentage (x 100 dans la Figure 6.1) pour rendre les données plus lisibles en termes de proportion d'emploi des variantes.

recherche. En effet, s'ils ont fait leurs preuves dans de nombreuses recherches (Ferret, *et al.*, 2001 ; Morin, *et al.*, 2004, etc.), nous avons dans notre cas été confrontée à divers problèmes de maintenance et à divers problèmes techniques de compatibilité entre versions et logiciels qui étaient tels qu'une sélection manuelle s'est avérée plus opérationnelle. C'est donc manuellement que nous avons repéré les variantes dans nos corpus, à l'aide du concordancier AntConc, de l'outil Syntex et de programmes Perl *ad hoc* (Chapitre3). Dans certains cas, les variantes nous ont été communiquées par les experts, en particulier certains cas de synonymie plus difficiles à repérer. Il s'agit donc d'un repérage outillé, mais non automatique.

TTVS		Ensemble des variantes repérées	Variantes sélectionnées (empreintes)
« Variante source » /Nombre de concepts concernés		84	60
Variation par changements graphiques	orthographiques	23	22
	siglaison	30	26
	flexion	2	3
Variation par changements morpho-syntaxiques	Changements qui ne modifient pas la structure du terme	3	1
	Changements entraînant une modification de la structure	15	5
Changement par réduction		26	22
Variation par changements lexicaux		6	11
Variation par changements complexes		11	4
Traductions		12	11
Nombre total de variantes		212	165

Tableau 6.4 – Sélection des variantes (TTVS)

DORIS		Ensemble des variantes repérées	Variantes sélectionnées (empreintes)
« Variante source » /Nombre de concepts concernés		37	25
Variation par changements graphiques	orthographiques	0	0
	siglaison	17	12
	flexion	1	1
Variation par changements morpho-syntaxiques	Changements qui ne modifient pas la structure du terme	0	0
	Changements entraînant une modification de la structure	5	3
Changement par réduction		7	6
Variation par changements lexicaux		10	8
Variation par changements complexes		6	6
Traductions		4	0
Nombre total de variantes		87	61

Tableau 6.5 – Sélection des variantes (DORIS)

La variation terminologique est un phénomène non négligeable dans nos corpus et en diachronie : les tableaux ci-dessus montrent que 71% des concepts qui présentent des variantes dans le TTVS (60 sur 84) et 67% des concepts qui présentent des variantes dans DORIS (25 sur 37) peuvent être sélectionnés sur la base des empreintes de fréquence et font donc potentiellement l'objet d'évolution dans le domaine.

Notons aussi à titre de comparaison que Daille (2005 : 181) estime que les variantes représentent entre 15 et 35% de la terminologie du domaine. Dans notre étude cependant, les variantes observées représentent 8,5% des termes dans le TTVS et 10% dans DORIS. Cette proportion est donc plus basse que celle estimée par Daille, aspect certainement lié en partie au fait que nous ne tenons pas compte ici des variantes syntaxiques souvent privilégiées dans les perspectives TAL (dont les expansions). La proportion des variantes dans nos corpus reste cependant importante et éclaire sur la centralité de ce phénomène dans les terminologies et en diachronie.

Le tableau suivant indique pour chaque empreinte sélectionnée (apparition simultanée de variantes et apparition/disparition de certaines variantes) le nombre et la répartition des variantes concernées.

DORIS		Apparition simultanée		Apparition/ disparition de certaines variantes	
		TTVS	DORIS	TTVS	DORIS
« Variante source » /Nombre de concepts concernés		20	14	40	11
Variation changements graphiques	par orthographiques	4	0	18	0
	siglaison	7	3	19	9
	flexion	0	1	3	0
Variation changements morpho- syntactiques	par Changements qui ne modifient pas la structure du terme	0	0	1	0
	Changements entraînant une modification de la structure	2	2	3	1
Changement par réduction		8	1	14	5
Variation par changements lexicaux		3	4	8	4
Variation par changements complexes		0	1	4	5
Traductions		4	0	7	0
Nombre total de variantes		48	26	117	35

Tableau 6.6 – Répartition des variantes sélectionnées en diachronie (TTVS et DORIS)

Il est difficile de tirer des conclusions générales au niveau de chaque type de variante à partir de ce tableau. En effet, nos corpus sont petits et malgré la proportion intéressante de variantes à observer, elles ne comptent que trop peu d'occurrences pour repérer des regroupements très marqués en fonction de leur type. De plus, rappelons qu'aucune évolution ne peut être prédite à partir des différents types de variantes.

Ceci dit, plusieurs observations peuvent ressortir. La première est le nombre marqué de variantes orthographiques dans le TTVS, en particulier en comparaison avec DORIS. Dans ces cas, si l'une des variantes finit par s'implanter dans le corpus, ceci peut être dû à la nature du corpus TTVS qui, dans la mesure où il est composé de trois éditions successives d'un même cours, reflète en partie les cas d'harmonisation orthographique et éditoriale d'une édition à l'autre.

Deuxièmement, dans les deux corpus la siglaison est très présente et intervient principalement dans les cas d'apparition/disparition de variantes. Ceci semble indiquer que dans le temps les termes ont tendance à être siglés.

Troisièmement, dans les deux corpus, le phénomène de réduction est également l'un des plus fréquents, notamment dans le TTVS. Nous verrons également qu'il s'agit du type de variantes qui permet d'illustrer l'opposition entre les variantes discursives et lexicales dont nous avons dit à la section 6.1.2 (p.193) qu'elles devaient être distinguées en diachronie. Dans le corpus DORIS, ce sont les changements lexicaux qui sont parmi les plus fréquents. Notons enfin que dans le corpus DORIS les termes/concepts concernés par les variantes sont assez équilibrés entre les deux empreintes (14 dont les variantes apparaissent de manière simultanée et 11 qui apparaissent/disparaissent) alors que dans le TTVS deux fois plus de termes/concepts sont impliqués dans des cas d'apparition/disparition de variantes. Encore une fois les deux corpus montrent donc des caractéristiques différentes pour un même indice.

Les experts sont interrogés sur l'évolution des concepts que ces variantes dénomment. Dans un premier temps nous détaillons la pertinence et les descriptions des évolutions liées aux variantes, puis nous revenons sur la distinction entre variantes discursives et variantes lexicales pour affiner l'interprétation.

6.3 Observation en corpus

6.3.1 Évaluation de la pertinence des variantes terminologiques en diachronie

Comme le montrent les tableaux suivants (Tableau 6.7 et Tableau 6.8), que ce soit dans le TTVS ou dans DORIS, la majorité des concepts pour lesquels plusieurs variantes existent sont jugés pertinents par les experts du point de vue de l'évolution : 42 sur 60 dans le TTVS, soit 70% et 23 sur 29 dans DORIS, soit 79%.

Notons que dans 7 cas (6 dans le TTVS et 1 dans DORIS) les experts n'ont pas su juger si le concept en question était oui ou non pertinent du point de vue de l'évolution.

TTVS	Nombre de concepts	Concepts dont les variantes sont apparues simultanément	Concepts dont certaines variantes apparaissent/disparaissent
Pertinents	42	12	30
Non pertinents	12	6	6
Ne se prononcent pas	6	2	4
Total	60	20	40

Tableau 6.7 – Pertinence des variantes sélectionnées en diachronie (TTVS)

DORIS	Nombre de concepts	Concepts dont les variantes sont apparues simultanément	Concepts dont certaines variantes apparaissent/disparaissent
Pertinents	23	12	11
Non pertinents	5	2	3
Ne se prononcent pas	1	0	1
Total	29	14	15

Tableau 6.8 – Pertinence des variantes sélectionnées en diachronie (DORIS)

Parmi l'ensemble des termes/concepts et leurs variantes pour lesquels on observe une évolution, des tendances très différentes se dégagent en fonction de l'empreinte de fréquence qui leur est associée. C'est ce que nous détaillons dans les deux paragraphes suivants.

6.3.2 Apparition simultanée de variantes

Que ce soit dans le corpus DORIS ou dans le corpus TTVS, le fait qu'un concept et ses variantes apparaissent simultanément dans les sous-corpus les plus récents confirme des évolutions déjà décrites lors de l'analyse des empreintes de fréquence. Les premiers phénomènes renvoient en effet à des observations sur la nouveauté de certains termes/concepts ou l'apparition de nouvelles fonctionnalités. Ainsi, dans le TTVS *mini-satellite**minisatellite* sont deux variantes pour un concept nouveau au début des années 90. On notera que ces deux variantes apparaissent simultanément, mais que dès 2002, la variante *minisatellite* est la seule utilisée dans le corpus. Cet exemple illustre le fait que lorsque des variantes graphiques sont concurrentes, la plus simple tend à s'implanter. Il

s'agit en effet d'une tendance marquée dans le TTVS, dont nous aurons l'occasion de montrer d'autres exemples dans notre travail. Dans le corpus DORIS, comme nous l'avons vu, la nouveauté est relative à un domaine d'activité et nous avons donc choisi de parler de l'apparition d'une nouvelle fonctionnalité. Ici, on retrouve la fonctionnalité *visibilité de satellite* et sa variante *visibilité satellite*.

Dans d'autres cas, les experts ont préféré parler de « récence » des concepts proposés plutôt que de nouveauté. En effet, dans le TTVS, pour certaines variantes telles que *PAO*|*polissage assisté par ordinateur*, *fibre optique dopée*|*fibre dopée*, *LDTP*|*Laser Diode Transmitter Package*|*source laser de communication*, les concepts en question sont apparus quelques années avant de faire leur entrée dans le cours et sont jugés plutôt comme des concepts récents. On retrouve dans ces variantes plusieurs termes/concepts liés à la question des lasers, dont nous avons dit qu'ils étaient au centre des recherches en optique spatiale depuis la fin des années 80, mais aussi la question des miroirs, à travers la thématique du polissage. Nous avons dit que la question des miroirs réapparaissait dans le corpus TTVS dans la mesure où de nombreux progrès les concernent. Ces progrès sont notamment dus au développement du *polissage assisté par ordinateur* que les experts notent comme récent en 1998. Le fait que les experts préfèrent parler de *récence* plutôt que de *nouveauté* fait écho à l'aspect mentionné au Chapitre 3 sur le fait que le domaine spatial est un domaine « à double vitesse » dont nombre de concepts ont d'abord été développés en laboratoire avant d'être intégrés pleinement au spatial. De ce fait, le jugement des experts est rendu plus difficile et peut les faire hésiter entre récence et nouveauté. Chacun de ces cas de figure est cependant légèrement différent du foisonnement terminologique ou de la synonymie de concurrence décrits entre autres par Dury ou Guilbert (il s'agit néanmoins bien de concurrence).

La concurrence entre ces variantes doit être mise en lien avec le rappel de thématiques anciennes telles que celle des miroirs, ce qui transparaît à travers l'apparition de variantes telles que *décentrement du miroir concave*|*Décentrement Miroir Concave*. Dans ce cas, à l'inverse du terme/concept *polissage assisté par ordinateur* et de sa variante *PAO*, il ne s'agit pas de termes/concepts récents. Ceci fait écho à plusieurs remarques établies aux chapitres 4 et 5 sur le fait que l'évolution des connaissances implique de « rappeler » des termes/concepts

anciens pour décrire des changements (§4.3.2.1, p.132 et §5.3.3, p.182). Pour préciser ce phénomène, on peut faire appel à la notion de *centralité*. Cette notion est utilisée notamment par Jacques (2005) qui montre dans son étude que l'utilisation de termes réduits dans un domaine est révélatrice de la centralité du référent dans le domaine, c'est-à-dire du « fait que, dans une sphère d'activités donnée, certains référents sont plus au centre des préoccupations des locuteurs de cette sphère d'activité que d'autres référents » (*ibidem* : 306). Dans notre cas, à l'inverse de Jacques, nous n'observons pas la réduction de variantes comme « conséquence » de la centralité d'un « référent », mais adoptons la notion de centralité pour décrire le fait qu'une thématique nouvelle, importante et discutée à un moment donné dans le domaine implique au niveau terminologique le rappel et l'utilisation de certains termes connus et de leurs variantes. En effet, dans le cas des variantes *décentrement du miroir concave* | *Décentrement Miroir Concave*, leur apparition est liée à la centralité de la thématique des miroirs à partir du milieu des années 90 dans le domaine.

Dans le corpus DORIS, les phénomènes d'évolution observables sont beaucoup plus homogènes et renvoient une fois encore à la migration des termes/concepts de l'informatique dans le projet. Ainsi, on repère des variantes telles que *bloc de télégestion* | *bloc TLG*, *émission des blocs de téléchargement* | *émission des TCHs*, *télégestion logicielle* | *TLGL* qui n'apparaissent que dans la troisième génération de l'instrument DORIS. L'expert interrogé a choisi lui aussi dans ce cas de parler de *récence* de ces concepts. Ceci peut s'expliquer par le fait que ces termes/concepts sont apparus progressivement entre 1990 et 2000, intervalle assez long pour que les experts n'aient plus le sentiment qu'ils soient nouveaux. Ceci peut néanmoins être discuté et l'expert lui-même parle parfois de nouveauté.⁹⁵

6.3.3 Apparition/disparition de certaines variantes

Dans le cas de l'apparition/disparition de certaines variantes dans les sous-corpus, les tendances d'évolution repérables sont hétérogènes. Au moins 5 cas de figure peuvent être identifiés.

⁹⁵ Nous reviendrons sur la question de l'instabilité du jugement expert dans la troisième partie de ce manuscrit.

Le premier cas, qui est un exemple intéressant mais assez difficilement repérable en diachronie courte, est celui du terme/concept *synthèse d'ouverture optique* dans le TTVS. En effet, comme le Tableau 6.9 l'indique, le calcul du coefficient d'implantation que l'on peut appliquer à ce type de cas (cf. *supra*) montre que la variante *synthèse d'ouverture* est la variante implantée à partir de 2002, parmi les trois variantes existantes. On observe donc sur un très court intervalle une concurrence de variantes suivie par une implantation complète de l'une d'entre elles. Dans ce cas, en plus de la nouvelle dénomination implantée (de 33% des occurrences en 1994, le terme *synthèse d'ouverture* en représente 100% en 2002), ce terme renvoie bien à un concept récent et prometteur au début des années 1990, sur lequel le Cnes travaille fortement. Avec le temps, ce concept est devenu de plus en plus familier dans le domaine et aujourd'hui, on engage toujours beaucoup d'efforts sur cette question. L'implantation d'une dénomination est ici parallèle à l'implantation du concept dans le domaine, ce qui illustre un cas exemplaire de parallélisme entre évolution de la dénomination (dimension interne) et du concept (dimension externe).

Variantes	Coefficient d'implantation		
	TTVS1994	TTVS1998	TTVS2002
<i>synthèse d'ouverture optique</i>	41,67	40	0
<i>synthèse d'ouverture</i>	33,33	60	100
SO	25	0	0

Tableau 6.9 – Coefficient d'implantation des variantes de *synthèse d'ouverture optique* (TTVS)

La deuxième tendance est liée à la précédente, mais elle ne permet pas de repérer un cycle complet d'implantation terminologique. En effet, dans un exemple tel que celui présenté au Tableau 6.10, on voit le processus d'implantation d'une variante terminologique : à partir de la dénomination *liaisons entre satellites* privilégiée en 1994, deux nouvelles dénominations concurrentes émergent (*liaison inter-satellites* et *liaison intersatellite*) dont on peut faire l'hypothèse que dans les prochaines années elles seront les variantes implantées. En particulier, il semble probable que *liaison intersatellite* soit aujourd'hui la variable implantée dans le domaine. Néanmoins, cet exemple n'illustre pas un cycle complet dans la mesure où dans l'intervalle de 8 ans du TTVS, on ne peut qu'observer le processus en train de s'accomplir. De plus, selon les experts, le concept est très bien implanté dans le domaine

depuis le début des années 1990 et est de plus en plus mis en avant *via* la mise en place de projets (tels que le projet Silex).

	Coefficient d'implantation		
Variantes	TTVS1994	TTVS1998	TTVS2002
<i>liaisons entre satellites</i>	75	44,44	40
<i>liaison inter-satellites</i>	0	0	10
<i>liaison intersatellite</i>	0	33,33	30
<i>liaisons LEO-GEO</i>	25	22,22	20

Tableau 6.10 – Coefficient d'implantation des variantes de liaisons entre satellites (TTVS)

Le troisième cas regroupe des variantes dont l'apparition/disparition indique une évolution des connaissances, mais non pas l'implantation d'une nouvelle dénomination. Dans ce cas, l'évolution des variantes doit être liée à la centralité de certaines thématiques dans le corpus à une période précise. Ce cas peut être illustré par les variantes liées à la thématique de l'infrarouge (Tableau 6.11) :

	Coefficient d'implantation		
Variantes	TTVS1994	TTVS1998	TTVS2002
<i>IR</i>	0	30,43	24,14
<i>Infrarouge</i>	97,22	69,57	75,86
<i>Infra-rouge</i>	2,78	0	0
<i>IRT</i>	0	55,56	43,48
<i>Infrarouge Thermique</i>	100	44,44	56,52
<i>MIR</i>	25	66,67	80
<i>Moyen-infrarouge</i>	0	33,33	20
<i>Moyen Infrarouge</i>	75	0	0
<i>PIR</i>	0	56,25	52,63
<i>Proche-infrarouge</i>	0	0	5,26
<i>Proche Infrarouge</i>	100	43,75	42,11
<i>VIS</i>	0	55,56	41,38
<i>visible</i>	100	44,44	58,62

Tableau 6.11 – Coefficient d'implantation des variantes liées à l'infrarouge (TTVS)

On le voit dans ce tableau, pour chacun des termes *visible*, *infrarouge*, *infrarouge thermique*, *proche infrarouge*, des variantes apparaissent à partir de 1998, et majoritairement des sigles. On notera l'apparition de la variante *VIS* pour *visible* qui n'est pas un sigle, mais une

truncation du terme *visible* mis en majuscules certainement par effet d'imitation des autres sigles.

Dans cet exemple, ce qui est remarquable est le fait que ces concepts appartiennent à un même paradigme⁹⁶ et que tous subissent la même tendance d'apparition/disparition de certaines variantes. Or, aucun de ces concepts et aucune des variantes n'est nouveau dans le domaine ou sa terminologie. L'apparition de ces variantes est liée à la centralité de la thématique de l'infrarouge à partir du milieu des années 90 dans le domaine. Les variantes existantes pour ces termes/concepts sont donc connues avant 1998, mais elles accompagnent le développement de la thématique de l'infrarouge dans le cours. En effet, cette thématique est récurrente en optique spatiale mais elle est au centre des préoccupations au milieu des années 90 grâce à des progrès tels que ceux des techniques de refroidissement (§9.1.2, p.279), mais également grâce à la mise en place de projets spatiaux tels que le projet IASI (Interféromètre Atmosphérique de Sondage dans l'Infrarouge). Dans la mesure où cette thématique est au centre des préoccupations du Cnes, on en parle plus dans les textes et l'on a besoin de diversifier les dénominations de ces concepts pour s'adapter à différents contextes et situations discursives. Cet élément introduit une distinction importante quant au statut des variantes à observer et à la distinction développée au début de ce chapitre entre variantes lexicales et variantes discursives.

La quatrième tendance que l'on peut observer est le cas de la « familiarisation » de concepts dans le domaine. En effet, lorsqu'un concept devient de plus en plus familier des experts, il est fréquent d'observer une simplification orthographique des dénominations. Ces cas sont très fréquents, que ce soit dans DORIS ou dans le TTVS, et se restreignent majoritairement aux variantes orthographiques et aux variantes obtenues par des changements morphosyntaxiques. Par exemple :

bi-CMOS → *BiCMOS* (TTVS)

palier de référence → *palier référence* (DORIS)

balise maîtresse DORIS → *balise maîtresse* → *BM* (DORIS)

oscillateur ultra-stable → *OUS* (DORIS)

coffret de batterie → *coffret batterie* (DORIS)

⁹⁶ Paradigme que l'on peut identifier à travers la notion de « dépendances syntaxiques » développée au chapitre suivant.

Il s'agit de cas où il est possible de corrélér assez significativement un type de variantes (les variantes graphiques et les variantes par changement morphosyntaxique) et un type d'évolution (évolution de la dénomination d'un concept et « familiarisation » de ce concept dans le domaine). L'idée de « familiarisation » d'un concept dans le domaine signifie que ce dernier est parfaitement intégré et stabilisé dans les connaissances de ce domaine et qu'il a vraisemblablement fini son implantation. Autrement dit, le parallélisme entre dimension interne et dimension externe au niveau de la relation terme/concept est illustré ici, puisque l'on pourrait décrire ce phénomène comme le parallèle entre une stabilisation graphique de la dénomination et la stabilisation du concept dans le domaine. Cet aspect est également marqué par des exemples de variantes « flexionnelles ». C'est l'exemple du terme/concept ASIC dans le TTVS qui apparaît au pluriel sous une forme invariable en 1994 (*des ASIC*) mais marque la flexion avec un -s graphique dès 1998 (*des ASICs*). Ce cas semble également indiquer une plus forte intégration de ce terme/concept dans le domaine, comme le confirment les experts. Néanmoins, il faut souligner que cette flexion peut aussi avoir pour cause l'influence de la langue anglaise dans le domaine, langue où l'on marque graphiquement la flexion du pluriel des sigles.

Enfin, le cinquième et dernier cas que l'on peut repérer lors de l'observation de variantes qui apparaissent/disparaissent est comparable au précédent, mais il s'agit d'un changement de dénomination planifié par une commission compétente. En l'occurrence, dans le cas de DORIS, il s'agit du service Contrôle-Qualité du Cnes qui, entre la première et la troisième génération de l'instrument DORIS, a pris la décision de recommander le terme *anomalie* à la place du terme *panne* (Tableau 6.12).

Variantes	Coefficient d'implantation	
	DORISGen1	DORISGen3
<i>panne</i>	95,34	4,66
<i>anomalie</i>	0	100

Tableau 6.12 – Coefficient d'implantation des variantes de *panne* (DORIS)

Ce service a en effet jugé que le terme *panne* revêtait une connotation négative qui pouvait faire hésiter les autres opérateurs à rapporter la panne, dans la mesure où ces derniers pouvaient avoir le sentiment de porter un jugement négatif sur le travail de leurs collègues.

6.3.4 Variantes lexicales et variantes discursives : recours aux contextes pour l'analyse

Pour terminer, nous revenons sur l'une des particularités du traitement des variantes pour la diachronie, liée au fait que ce phénomène est susceptible de relever du niveau discursif ou du niveau lexical. Comme illustré dans ce chapitre, ces deux niveaux sont pertinents pour observer l'évolution et, selon le niveau sur lequel les variantes interviennent, les interprétations d'évolution peuvent différer.

Dans le cas où les variantes repérées interviennent au niveau lexical, ceci entraîne deux types d'interprétations qui peuvent être complémentaires :

1. ces variantes marquent la création d'une nouvelle dénomination, complétée ou en cours. En d'autres termes, l'apparition/disparition de variantes en corpus indique un changement terminologique stable et lexicalisé ou en cours de lexicalisation :
 - a. soit la nouvelle dénomination est implantée (par exemple *panne* | *anomalie*)
 - b. soit la nouvelle dénomination est en cours d'implantation (par exemple *liaisons entre satellites* | *liaison intersatellite*)
2. ces variantes marquent la familiarité ou l'implantation du concept dénommé dans le domaine (par exemple *synthèse d'ouverture optique* | *synthèse d'ouverture*).

Dans le cas où les variantes interviennent au niveau discursif, un seul type d'interprétation est possible : les variantes marquent simplement la centralité du concept qu'elles dénomment à une période précise du domaine (par exemple la thématique de l'infrarouge), sans indiquer de changement terminologique stable.

Ceci implique dans le cas des « variantes lexicales » que deux niveaux d'évolution des connaissances sont en jeu : l'évolution du concept dénommé et l'évolution de la

dénomination. Dans le cas où les variantes sont « discursives », un seul niveau d'évolution est en jeu, celui du concept.

Cependant, comme décrit entre autres par Alber-Dewolf (1982) ou Jacques (2003), ces deux niveaux sont perméables et avant d'être lexicalisée, une variante est souvent d'abord discursive⁹⁷. Ces deux auteures montrent que c'est notamment ce qui se passe dans le cas de la naissance d'une variante réduite, qui dans un premier temps a été une variante discursive d'une variante longue. Le temps aidant, et le concept devenant de plus en plus central dans le domaine (Jacques, 2005), c'est la variante réduite qui s'implante. Bien que ce type de passage entre niveau discursif et niveau lexical soit difficilement observable en diachronie courte, dans notre corpus, le cas de *synthèse d'ouverture optique* | *synthèse d'ouverture* en est un exemple.

Soulignons également que la création d'une variante à caractère lexical ne signifie pas que celle-ci ne peut plus intervenir sur le plan discursif. Plus précisément, dans l'exemple de *synthèse d'ouverture*, bien que cette variante soit maintenant implantée et soit la nouvelle dénomination préférée dans le domaine pour ce terme/concept, il n'est pas impossible que l'on retrouve cette variante en concurrence avec la variante longue *synthèse d'ouverture optique* en discours. La variante réduite, bien que lexicalisée, peut alors dans certains cas être encore la reprise anaphorique de la variante longue. En effet, la distinction que nous faisons ici entre variante lexicale et variante discursive n'implique pas une exclusion totale de ces deux niveaux, mais simplement une piste d'analyse pour différencier des évolutions telles que la centralité d'un terme/concept et l'implantation d'un concept récent dans le domaine. Cette question mériterait un développement important qui dépasse les limites de notre étude, mais nous renvoyons le lecteur vers des travaux tels que ceux de Collet (1996) et Jacques (2003) pour étayer ces pistes de réflexion.

D'un point de vue méthodologique, pour repérer si la variante intervient au niveau lexical ou au niveau discursif, une piste possible est de compléter l'analyse à travers l'observation

⁹⁷ Jacques (2003 : 70) insiste sur l'importance de la dimension discursive pour expliquer la création néologique par réduction, et critique le manque de travaux expliquant les mécanismes discursifs en jeu. Nous n'apportons pas de réelle réponse à cette question dans notre travail, mais nous nous accordons avec l'auteure et insistons sur l'importance des descriptions des mécanismes de variation en discours pour l'analyse diachronique.

des contextes dans lesquels elles apparaissent. Par exemple, deux types de contextes sont observables dans nos corpus : des contextes « équivalents » et des contextes « complémentaires ». Dans le premier cas, les variantes observables relèvent le plus souvent du niveau lexical et dans le second du niveau discursif.

Par exemple dans DORIS, les variantes *Ampli* | *amplificateur* peuvent être considérées comme des variantes qui interviennent uniquement au niveau discursif dans la mesure où ces deux variantes n'apparaissent jamais dans des contextes équivalents, mais plutôt dans des contextes « complémentaires » tels que des contextes phrastiques vs. des contextes particuliers tels que des figures, des symboles et des tableaux. Par exemple dans DORIS :

- contexte phrastique :

- (lxviii) ... Les dissipateurs des **amplificateurs** seront en face arrière du coffret...
 ... entreprises sont de mettre hors tension les **amplificateurs** de puissance, le synthétiseur et la carte météo...

- liste (tableau) :

- (lxix) DORIS : Liste des connecteurs du coffret d'émission et de gestion :
- | | |
|-----|---|
| J34 | 9 points Ampli 400 MHz |
| J9 | Liaison CDC |
| J35 | SMA sortie 400 MHz Ampli 400 MHz |
| J36 | 9 points Ampli 2 GHz |
| J11 | Connecteur DB25 face avant |
| J37 | SMA entrée 2 GHz Ampli 2 GHz |

Dans la mesure où ces variantes se partagent des contextes « complémentaires », on envisage difficilement que la variante *Ampli* vienne remplacer *amplificateur*.

Ce cas se retrouve également dans l'exemple de la thématique de l'infrarouge du TTVS où les sigles apparus (*IRT*, *VIS*, *MIR*, *IR*, *PIR*) marquent une tendance forte à apparaître dans des contextes particuliers tels que des tableaux et des formules alors que les formes longues apparaissent de manière privilégiée dans les contextes phrastiques. Par exemple, dans le TTVS1998 :

- contexte phrastique

(lxx) Pour éviter l'**infrarouge thermique** (au delà de 3 microns) et toute l'instrumentation associée qui est très lourde et peu efficace, on cherche à travailler dans le **proche infrarouge** (entre 1.3 et 1.7 microns environ).

- tableau

(lxxi)

<i>Mission</i>	<i>imagerie à haute résolution</i>
<i>Instrument</i>	<i>caméra</i>
<i>Type d'orbite</i>	<i>héliosynchrone</i>
<i>Champ/résolution</i>	<i>500 à 800 km</i>
<i>Domaine spectral</i>	<i>10 à 100 nm</i>
<i>Résolution</i>	<i>1 à 10m</i>
<i>Effet physique utilisé</i>	VIS+PIR(+IRT)
<i>Radiométrie</i>	<i>bandes larges : (10nm = VIS+PIR)</i>

Dans le cas des variantes *liaison entre satellites* | *liaison intersatellite* par contre, celles-ci apparaissent dans des contextes phrastiques très similaires. La seconde semble plutôt se substituer à la première dans des contextes équivalents, ce qui semble indiquer que la variante *liaison intersatellite* vient remplacer petit à petit la variante *liaison entre satellites* :

(lxxii) ... Dans le cas d'une **liaison entre satellites**, il existe encore deux classes d'applications ...
... LES APPLICATIONS On peut classer les **liaisons entre satellites** en trois grandes catégories ...

(lxxiii) ... laser est fournie par les **liaisons intersatellites** dans les constellations en orbite basse (LEO, ...
... 450 km Les 3 premières nécessitent des **liaisons intersatellites**, qui sont à 23 Ghz pour Iridium, et à 60 Ghz ou ...

Bien entendu, les phénomènes que nous décrivons ne sont que des tendances sur lesquelles l'analyste peut s'appuyer pour tenter de caractériser au mieux les types de variantes en jeu dans l'évolution et il est fréquent que la répartition entre types de contextes ne soit pas aussi nette (par exemple dans le cas de la thématique de l'infrarouge, certains sigles apparaissent dans des contextes phrastiques, bien que cela soit moins fréquent que dans le cas des variantes longues). De plus, d'autres éléments peuvent (voire doivent) venir compléter les types d'indices que nous décrivons.

Par exemple, dans le cas où une des variantes est une réduction, pour distinguer les variantes discursives, on peut aussi rechercher en contexte la présence de la variante longue comme antécédent. Si la variante réduite est précédée en discours de la variante longue, il s'agit d'une variante discursive⁹⁸ :

- (lxxiv) *La figure VII.79 (A) montre que, même si l'instrument possède une pupille de sortie refroidie, il n'est pas possible d'éviter le **flux lumineux parasite** émis par l'environnement car le champ de l'instrument est trop important : le détecteur central "a" ne voit que l'objet à travers les optiques. [...] Ce **flux parasite** peut être du même ordre de grandeur que le signal utile voire supérieur.*

Dans ce contexte, le lien entre la variante « antécédent » et la variante réduite est également marqué par la présence du démonstratif qui souligne la reprise anaphorique et donc le caractère discursif de cette variante.

Dans cet exemple, cet aspect est d'autant plus important qu'il vient justifier une première observation « contre-intuitive ». En effet, dans le cas des variantes *flux lumineux parasite* et *flux parasite*, la variante la plus courte est présente dans les sous-corpus les plus anciens du TTVS, et la variante la plus longue n'apparaît qu'à partir de 1998. D'un point de vue diachronique, ceci est surprenant dans la mesure où habituellement la langue a recours à différents procédés de réduction ou de simplification des termes pour dénommer un même concept (et non pas à l'expansion pour dénommer un concept équivalent). Dans cet exemple, ce phénomène s'explique alors du fait que l'on se trouve face à des variantes discursives qui indiquent la centralité du concept dans le domaine à une période donnée.

Ces différents aspects, qui reposent sur la distinction entre niveau lexical et niveau discursif, ne sont que des pistes de réflexion dans cette recherche, mais ils permettent de souligner l'importance de prendre en compte l'analyse des variantes en discours et non pas au seul niveau lexical. Les variantes en diachronie ne doivent pas être considérées comme des indices de néologie seulement, mais également comme des indices de centralité où le niveau discursif entre en jeu. De plus, bien qu'il soit tentant de chercher à observer à l'aide des variantes un parallèle régulier entre une évolution de la dénomination et une évolution du

⁹⁸ Bien que, comme le souligne Jacques (2005 : 198*sqq.*), il est possible qu'une variante discursive se réalise sans antécédent proche.

concept, ce point de vue serait réducteur face à la complexité des phénomènes d'évolution des connaissances liés à l'analyse des variantes⁹⁹.

6.4 Bilan

Les variantes terminologiques comme indices d'évolution des connaissances d'un domaine se révèlent elles aussi très riches d'informations. Les évolutions qu'elles permettent de repérer sont multiples et hétérogènes et font écho à certaines analyses menées à l'aide des deux indices précédents.

D'un point de vue méthodologique, cet indice n'est pas toujours évident à mettre en œuvre et ce, pour plusieurs raisons. Tout d'abord, il n'est pas toujours facile pour les experts de juger de la pertinence ou non des concepts concernés par les variantes. De plus, préciser le type d'évolution en jeu n'est pas toujours aisé en raison de la faible quantité de données et du faible nombre d'occurrences de certaines variantes. Néanmoins, la proportion de variantes pertinentes pour la diachronie reste très intéressante et le « potentiel » de cet indice peut être développé par d'autres stratégies d'analyse pour aiguiller la recherche, telles que l'observation des contextes et la distinction entre variantes lexicales et variantes discursives.

Néanmoins, il reste que lorsque l'on entame une analyse de l'évolution des connaissances en corpus, il peut être intéressant de commencer l'exploration par des indices plus généraux et globaux tels que les empreintes de fréquence et les contextes riches en connaissances évolutives qui sont plus faciles d'accès et d'interprétation pour l'analyste. De plus, les variantes terminologiques doivent être liées étroitement aux empreintes de fréquence pour pouvoir être observées d'un point de vue diachronique. Cet indice peut donc apparaître plus comme un « renfort » à l'analyse des empreintes de fréquence que comme un indice isolé à observer. Nous détaillerons cet aspect au Chapitre 9, mais il est clair à ce stade que des indices tels que les variantes sont d'autant plus pertinents qu'ils peuvent se combiner à d'autres.

⁹⁹ Ce que montrent également les résultats mitigés présentés par Tartier (2004) dont l'hypothèse repose essentiellement sur le parallèle entre évolution du concept et évolution de la dénomination.

Dans ce chapitre, nous avons montré également que bien que les variantes terminologiques invitent à poser un parallèle entre l'évolution de la dénomination et l'évolution du concept, l'observation de l'évolution doit dépasser ce niveau strictement terminologique. En effet, il est non seulement difficile de faire émerger des corrélations entre types de variantes¹⁰⁰ et types d'évolution, mais il est également réducteur de considérer les variantes en dehors des textes et des discours. L'intérêt des variantes en diachronie repose autant sur le fait que la dénomination peut évoluer que sur le fait qu'il s'agit d'un signe linguistique dont le comportement en discours est également un indice d'évolution.

De plus, l'analyse des variantes montre l'indépendance relative qui peut exister entre l'évolution des dénominations et l'évolution des concepts : la dénomination peut évoluer sans entraîner de modification au niveau conceptuel et vice versa.

Pour compléter ces observations, dans l'avant-dernier chapitre de cette partie, nous décrivons un phénomène linguistique que certaines études classent dans le champ des variantes terminologiques, mais qui permet plutôt d'aborder la question de la cooccurrence en diachronie : les dépendances syntaxiques.

¹⁰⁰ Si ce n'est une légère tendance des variantes graphiques à pouvoir être corrélées à la familiarisation/stabilisation d'un terme/concept dans le domaine.

Chapitre 7 Dépendances syntaxiques

SOMMAIRE DU CHAPITRE

7.1	DÉPENDANCES SYNTAXIQUES ET ANALYSE DIACHRONIQUE	221
7.1.1	<i>Dépendances et cooccurrences : définitions</i>	221
7.1.2	<i>Exploration des dépendances en diachronie</i>	224
7.2	EXTRACTION DES DÉPENDANCES SYNTAXIQUES EN CORPUS	231
7.2.1	<i>Analyse syntaxique automatique de l'outil Syntex</i>	231
7.2.2	<i>Sélection des dépendances</i>	232
7.3	OBSERVATIONS EN CORPUS	236
7.3.1	<i>Évaluation de la pertinence globale des dépendances syntaxiques en diachronie</i>	236
7.3.2	<i>Nouveauté et obsolescence : productivité morphologique et lexicale</i>	238
7.3.3	<i>Apparition/disparition de thématiques</i>	240
7.3.4	<i>Changement de point de vue</i>	243
7.4	BILAN	249

Dans ce chapitre, nous présentons un dernier indice linguistique pour repérer l'évolution des connaissances en corpus, indice qui repose sur l'hypothèse qu'il est possible de repérer des informations sur l'évolution du domaine en observant les évolutions de la cooccurrence des termes. Nous abordons ici cette notion sous la forme des dépendances syntaxiques.

Dans la première section, nous délimitons les dépendances par rapport à la notion de cooccurrence et décrivons l'apport de cet indice à travers différentes études diachroniques existantes. Dans la deuxième section, nous décrivons comment l'outil Syntex repère les dépendances et terminons par la description de trois phénomènes d'évolution que cet indice, riche d'informations, permet de traiter.

7.1 Dépendances syntaxiques et analyse diachronique

7.1.1 Dépendances et cooccurrences : définitions

La notion de cooccurrence est une notion centrale en linguistique, et notamment en Traitement Automatique des Langues, dans la mesure où elle permet de caractériser le comportement syntaxique et sémantique des unités lexicales. Firth (1957) est l'un des premiers à soumettre cette idée, comme l'illustrent ses mots devenus célèbres : « You should

know a word by the company it keeps. » (*ibidem* : 179). Cette idée se retrouve également dans les principes de l'analyse distributionnelle proposée originellement par Harris (1954 ; 1988), et dont l'hypothèse sous-jacente est que les mots de sens proche peuvent être caractérisés par leur apparition dans des contextes similaires. Selon cette hypothèse, l'étude de la cooccurrence permet d'induire le fonctionnement sémantique d'unités lexicales à partir de leurs contextes d'emploi : « characterizing words by their selection allows for considering the kind and degree of overlap, inclusion and difference between words in respect to their selection sets [...] » (Harris, 1991 : 332).

Dans la lignée des travaux de Harris, depuis une vingtaine d'années, l'analyse de la cooccurrence se trouve au centre de nombreuses recherches et analyses automatiques en corpus telles que l'acquisition de classes sémantiques (Grefenstette, 1994 Habert, *et al.*, 2004), le repérage de la polysémie (Bertels, 2005 ; 2006 ; Fabre, *et al.*, 1997) ou la désambiguïsation (Audibert, 2003 ; Hearst, 1991 ; Turney, 2004 ; Yarowsky, 1995). Mais bien que la notion de cooccurrence soit impliquée dans de nombreux travaux, dans les faits, et en fonction des besoins des analyses, cette notion répond à des réalités quelque peu différentes. En effet, la définition de la cooccurrence varie sur au moins deux aspects : la définition du *contexte* d'observation des cooccurrents et du *degré d'association* entre cooccurrents.

La notion de degré d'association renvoie au fait que les associations entre cooccurrents peuvent être plus ou moins libres. Ainsi, des travaux tels que ceux de Reymond (2002) ou Blumenthal (2008) soulignent le fait que la cooccurrence s'inscrit sur un continuum entre des associations libres de cooccurrents et des degrés de cohésion plus forts, tels que les collocations ou le figement. D'un point de vue automatique, ce degré de cohésion est généralement traduit en calculs statistiques (tels que l'entropie, l'information mutuelle, etc.) pour mesurer le degré d'association entre cooccurrents et ainsi identifier les cas de collocations et de figements (Kilgarrieff, 2006 ; Losee, 1998 : 43-62).

La notion de *contexte* pour l'observation des cooccurrents correspond à la « fenêtre » au sein de laquelle on sélectionne des cooccurrents potentiels pour une unité lexicale donnée. On

peut distinguer au moins trois points de vue différents sur la définition du contexte (Baroni & Evert, 2007) :

- les cooccurrences de surface : le contexte considéré est défini à l'aide d'une fenêtre de n mots et n mots après l'unité étudiée,
- les cooccurrences « textuelles » : les unités lexicales prises en compte gravitent dans la même phrase, le même paragraphe ou le même document que l'unité lexicale étudiée,
- et les cooccurrences syntaxiques : les cooccurents pris en compte entretiennent une relation de dépendance syntaxique avec l'unité lexicale étudiée.

Dans cette étude, c'est ce dernier point de vue qui sera considéré et c'est à travers la notion de *dépendances syntaxiques* que la question de la cooccurrence sera abordée en diachronie. Ce choix se justifie par au moins trois éléments.

Tout d'abord, d'un point de vue méthodologique et applicatif, les dépendances syntaxiques sont utilisées depuis plusieurs années dans le cadre de la construction de ressources terminologiques (par exemple par Aussenac-Gilles & Bourigault, 2003 ; Bouaud, *et al.*, 1997 ; Bourigault & Lamé, 2002). En effet, comme le notent par exemple Bourigault *et al.* (2004 : 14) les dépendances syntaxiques permettent d'élaborer des taxinomies locales dans la mesure où elles permettent de mettre en relation de dépendance un terme tête et une expansion. Cette relation fait émerger alors un lien sémantique entre ce terme/concept originel et le terme complexe issu de l'association [terme tête + expansion]. Le lien sémantique généralement mis au jour par les dépendances étant le lien d'hyponymie/hyperonymie, une des hypothèses en diachronie est que l'analyse des dépendances permet de faire émerger (entre autres) les termes/concepts plus spécifiques apparus dans le domaine.

Deuxièmement, l'analyse des dépendances syntaxiques permet d'automatiser les analyses sémantiques en corpus, même dans le cas de corpus de petite taille tels que les nôtres, à la manière de Habert et Nazarenko (1996) par exemple, qui travaillent sur un corpus médical d'un peu plus de 20 000 occurrences. Ceci est intéressant dans la mesure où généralement, les analyses sémantiques sur la base de cooccurrences requièrent de préférence de grands

corpus pour garantir la significativité des résultats de l'analyse distributionnelle (Grefenstette, 1994).

Enfin, la sélection de cooccurents sur un critère syntaxique nous semble moins arbitraire que la sélection *a priori* d'une fenêtre contextuelle, de la phrase ou du paragraphe comme zone d'analyse.

7.1.2 Exploration des dépendances en diachronie

Des travaux d'analyse des dépendances en diachronie ont été mis en place par Harris et son équipe dès la fin des années 80 (Harris, *et al.*, 1989). Ces travaux, résumés par Ryckman (1990 : 32-37), s'inscrivent dans le champ de l'analyse distributionnelle dont la prise en compte des dépendances syntaxiques constitue le cœur :

« [...] the structural property is not merely co-occurrence, or even frequent co-occurrence, but rather dependence on a word set: an operator does not appear in a sentence unless a word – one or another – of its argument set is there (or has been zeroed here). » (Harris, 1991 : 332)

Cette analyse diachronique s'attache à la description du « sous-langage » (pour respecter la terminologie de Harris) du domaine de l'immunologie. Plus précisément, l'objectif des chercheurs était de voir s'il était possible de « localiser et de caractériser les désaccords entre les chercheurs du domaine et plus généralement les changements d'information intervenant au cours du temps » (Ryckman, 1990 : 34) en analysant 14 articles scientifiques du domaine parus entre 1935 et 1966. Comme le décrit Ryckman (*ibidem* : 33-34), le domaine de l'immunologie a connu de grosses controverses dans les années 40 au sujet des cellules du système lymphatique, controverses dont Harris *et al.* font l'hypothèse qu'elles se reflètent dans l'évolution des structures informationnelles du sous-langage. Les auteurs analysent donc l'évolution des contraintes de combinaison des termes (c'est-à-dire leurs dépendances syntaxiques) dans l'objectif de définir ce qu'Harris appelle la « grammaire sémantique » du domaine, pour relever les éléments nouveaux entre 1935 et 1966. L'observation des combinaisons et structures des termes sur une trentaine d'années

« révèle plusieurs choses intéressantes du point de vue du domaine scientifique lui-même. Le changement continu de la connaissance apporté par de nouveaux tests immunologiques est indiqué par les changements intervenant dans les structures informatives des textes : apparition de nouvelles structures, disparition progressive ou rareté des anciennes. » (Ryckman, 1990 : 36)

Les changements mis au jour par cette analyse semblent très précis et mettent en évidence de véritables changements de paradigmes tels que Kuhn (1999) les a décrits et tels que les recherchent les auteurs :

« Les premiers articles contenaient presque exclusivement des séquences traitant des changement histologiques faisant suite à l'injection d'antigène. Les articles suivants mettent l'accent sur les différents types de cellules contenant des anticorps après l'injection répétée d'antigène. Après que le microscope électronique ait considérablement amélioré le seuil de résolution du visible, l'accent a été mis sur les changements cellulaires faisant suite à la stimulation antigénétique et on a déterminé que l'anticorps était présent dans des cellules particulières. » (*ibidem* : 36)

Ce type d'approche confirme le fait que l'analyse distributionnelle permet de retracer des évolutions de connaissances en diachronie, bien qu'aujourd'hui, à notre connaissance, très peu de travaux s'inscrivent dans ce type de démarche diachronique.

Aujourd'hui pourtant, bon nombre de recherches de TAL reprennent le principe de l'analyse distributionnelle harrissienne pour proposer des analyses sémantiques automatisées en corpus (par exemple Church, *et al.*, 1991 cités par Habert & Zweigenbaum, 2003, ou en France Fabre, *et al.*, 1997 ; Habert, *et al.*, 2004 ; Zweigenbaum & Habert, 2004¹⁰¹). Mais, comme le remarquent Habert et Zweigenbaum (2003 : 30), ces travaux ont plutôt recours à des méthodes « globalement distributionnelles » et ne mettent pas strictement en œuvre les principes de transformation et de normalisation des phrases proposés par Harris. Dans la mesure où ces travaux cherchent à automatiser l'analyse, le principe de l'analyse

¹⁰¹ On notera l'affiliation avec les travaux de Harris assumée par Habert *et al.*, dont l'outil d'analyse distributionnelle a été nommé Zellig, prénom de Harris.

distributionnelle alors essentiellement retenu est le fait que le « sens » des mots peut être caractérisé par leurs dépendances¹⁰².

L'objectif de ce type de recherches est de mettre au jour des classes sémantiques à partir de termes qui partagent l'ensemble ou une partie de leur distribution. Ce repérage est généralement divisé en trois étapes principales (Grefenstette, 1994) :

- l'extraction des cooccurents,
- le calcul (statistique) de distance entre mots sur la base des cooccurents qu'ils partagent,
- le regroupement des mots en classes en fonction de la distance calculée.

À notre connaissance, il n'existe pas encore d'extraction automatique de classes sémantiques en diachronie, bien que l'on voie l'intérêt qu'une comparaison de classes sémantiques dans le temps pourrait revêtir pour l'analyse de l'évolution. Ceci s'explique certainement en partie par la difficulté de l'extraction de telles classes, tant du point de vue de leur mise en œuvre que de l'interprétation des regroupements qui en sont issus. En effet, comme le soulignent notamment Habert et Nazarenko (1996 : 141) :

« les listes de mots obtenues ne constituent pas de véritables classes de mots cohérentes et homogènes. De surcroît la cohérence sémantique, ou le sens, des classes de mots résultantes n'est pas immanente, elle est le fruit d'une interprétation qui est d'autant plus difficile que les mesures de similarités constituent des résumés statistiques bruts dont les critères de construction sont effacés. »

De plus, on peut émettre l'hypothèse que ce type de démarche porte des résultats plus significatifs sur des intervalles temporels longs où l'évolution sémantique sera certainement plus marquée.

Il existe cependant d'autres types d'analyses sémantiques sur la base de l'observation des changements de cooccurrences d'un terme dans le temps, desquelles nous nous inspirons dans ce chapitre. Si l'on reprend les trois phases de l'extraction de classes proposées par

¹⁰² On retrouve ici le principe énoncé par Firth et présenté au début de ce chapitre.

Grefenstette *supra*, ce type d'analyse s'arrête à la première phase, celle de l'extraction des cooccurents. Il ne s'agit donc plus de faire émerger des classes de termes, mais d'observer la distribution de termes isolés pour en comprendre les évolutions. Ceci revêt un intérêt diachronique certain pour observer l'évolution sémantique de termes, dans la mesure où, comme le reformule Béjoint (1989 : 410) dans une perspective terminologique

« la complexité sémantique d'un lexème se manifeste par la quantité et la variété des collocations dans lesquelles il entre dans un corpus : plus un lexème est riche sémantiquement plus ses collocations sont nombreuses et variées [...]. Un lexème polysémique aura donc plus de collocations différentes qu'un lexème monosémique. »

De fait, on peut émettre l'hypothèse qu'une variation dans le temps de la cooccurrence d'un terme donné peut révéler une évolution sémantique de ce terme. Cette perspective est explorée par certains travaux, notamment en analyse de discours politiques où l'objectif n'est pas directement d'observer les changements de sens des termes, mais plutôt les changements de contextes d'emplois de ces termes, révélateurs d'une vision politique différente. De plus, ces analyses ne prennent généralement pas en compte les dépendances syntaxiques, mais plutôt des fenêtres contextuelles pour le calcul de cooccurents. C'est par exemple l'objectif de Leblanc et Martinez (2006) qui analysent les réseaux de cooccurrence du terme *monde* par les présidents de la Cinquième République.

Nous inscrivons plutôt notre recherche à la suite d'une proposition d'analyse de la cooccurrence dans le temps proposée par Condamines, Rebeyrolle et Soubeille (2004). Leur recherche prend place au Cnes dans le cadre de projets spatiaux et utilise une version différente du corpus DORIS choisi pour notre étude. Dans leur article, les auteures proposent de classer les termes en fonction des évolutions de leurs dépendances, qu'elles divisent en deux catégories : leurs expansions (modificateurs des termes têtes) et leur distribution (tous les autres contextes non pris en compte dans l'expansion, principalement les verbes dont les termes sont les arguments). Sur cette base, elles proposent quatre groupes

de termes, dont le premier regroupe les termes qui n'évoluent pas ou peu et le dernier les termes susceptibles d'évoluer le plus :

- termes dont ni l'expansion ni la distribution n'évolue
- termes dont l'expansion évolue mais dont la distribution reste stable
- termes dont l'expansion reste stable mais dont la distribution évolue
- termes dont l'expansion et la distribution évoluent simultanément.

L'exploration que les auteures proposent ne détaille malheureusement que très succinctement les termes concernés par cette analyse, mais leurs conclusions montrent le potentiel de ces regroupements sur la base des dépendances : le regroupement peut être outillé, mené assez aisément et révéler des phénomènes pertinents pour l'évolution des connaissances (cette dernière conclusion est étayée par le concours d'un expert lors de leur étude). De plus, dans le cas du corpus analysé, la majorité des termes observés présentent des distributions et des expansions différentes, ce qui laisse supposer que la variation dans le temps repérée à l'aide des dépendances est importante.

Dans la suite de ce chapitre, nous nous inscrivons dans la même veine que les auteures mais, à l'inverse de leurs travaux, nous ne distinguons pas les expansions de la distribution, et considérons simplement ces deux cas comme des dépendances syntaxiques à analyser¹⁰³. Ce choix est justifié par l'analyse de l'outil Syntex avec lequel sont extraites les dépendances et qui ne prend en compte aucune de ces distinctions dans son traitement. De fait, nous ne posons aucun *a priori* sur les types de dépendances syntaxiques à ce stade de l'analyse.

En plus de ces démarches d'inspiration harrissienne, au moins deux autres approches peuvent être mentionnées, qui analysent l'évolution des dépendances syntaxiques d'un point de vue différent. La première de ces approches concerne le repérage de néologismes, à travers l'extraction de nouvelles dépendances pour un terme donné. L'apparition de ces nouvelles dépendances a été décrite notamment par Ahmad *et al.* (2002) sous le terme de « productivité morphologique et lexicale », défini ainsi :

¹⁰³ Nous verrons cependant que les différents types de dépendances peuvent entrer en jeu. Notamment, les dépendances syntaxiques verbales fournissent des informations assez homogènes quant à l'évolution des connaissances, contrairement aux autres types de dépendances, plus hétérogènes (§7.3.4, p.243).

« When specialists write about a concept or an artefact, they start by describing one concept or artefact. Soon enough they find concepts or artefacts which they can relate to their original, and indeed, often form classes of concepts or artefacts. » (*ibidem* : 2)

Ceci signifie qu'un terme/concept peut donner naissance à un nouveau terme/concept auquel il est syntaxiquement ou morphologiquement lié, qui est plus spécifique et dont on peut retracer l'apparition en diachronie.

Cette idée peut être complétée par la notion d'*interface* proposée par Humbley (2006 : 93) :

« Si le néonyme représente un concept nouveau, il conserve un lien avec l'ancien, le connu, et ce lien est généralement représenté dans son expression linguistique, qui en constitue de fait l'interface. Prenons un exemple [...], un terme qui relève depuis peu de la vie de tous les jours : « centime d'euro ». L'apport néologique (de taille, car le référent voit sa valeur multipliée par 6,55957) est véhiculée par l'ajout de l'élément « euro », qui joue ainsi le rôle d'interface entre les deux monnaies – le franc, l'ancien, présent dans le terme consacré « centime », et l'euro, annoncé comme tel. »

La création de termes complexes par l'ajout d'une expansion constitue un cas où l'interface entre le terme/concept nouveau et l'ancien est facilement repérable :

« C'est dans le cas de nombreuses synapsies néologiques que l'interface rend les plus grands services en clarifiant les rapports entre l'état actuel des connaissances et la nouveauté apportée. » (*ibidem* : 98)

Pour reprendre les termes de l'auteur, l'idée d'interface revêt un intérêt « à la fois théorique et pratique » innovant qui permet de proposer des pistes pour un repérage de la néologie qui ne repose plus seulement sur le recours aux corpus lexicographiques d'exclusion, ni sur la description de matrices lexicogéniques seules (§1.2.3.2 et §4.1.2). Et de ce point de vue, l'analyse des dépendances syntaxiques en diachronie peut alimenter la réflexion autour de la notion d'interface, notamment dans une perspective méthodologique, *via* la description

des possibilités de repérer et « mesurer, d'après la forme du néologisme adopté, comment ou dans quelle mesure la nouveauté est signalée de façon explicite » (*ibidem* : 96).

La deuxième approche est celle de veille scientifique et technologique et est proposée par Ibekwe-Sanjuan et Sanjuan (1998 ; 2002). Ces travaux ont été brièvement introduits dans le chapitre précédent dans la mesure où les auteurs utilisent la notion de variantes de termes pour repérer des changements thématiques dans les textes. Dans les faits, les variantes qu'ils définissent relèvent de mécanismes tels que l'expansion de termes ou l'insertion, qui mettent en jeu les dépendances syntaxiques et qui impliquent un changement conceptuel entre les deux « variantes » obtenues. Le système de veille qu'ils proposent, TermWatch, trouve d'ailleurs son originalité dans l'utilisation de relations syntaxiques pour la veille au lieu d'autres critères de cooccurrences tels qu'une fenêtre de mots (Ibekwe-Sanjuan & Sanjuan, 2004). Pour ces raisons, et dans la mesure où nous avons défini les variantes comme entrant dans une relation d'équivalence conceptuelle, nous considérons que cette approche relève des dépendances syntaxiques.

Dans ces travaux, à partir de termes extraits, les auteurs repèrent des changements de dépendances syntaxiques qui affectent soit la tête du terme, soit ses expansions. Leur hypothèse est que ces modifications du terme indiquent l'apparition de nouvelles thématiques de recherches abordées dans les textes, ainsi que leur organisation dans des articles scientifiques. Plus précisément, le repérage de ces éléments permet aux auteurs de mettre au jour des classes conceptuelles et des relations génériques/spécifiques à partir desquelles le système TermWatch calcule l'organisation thématique des documents observés. À partir des changements qui affectent ces classes conceptuelles et les relations génériques/spécifiques, le système calcule les évolutions intervenant dans ces thématiques. Du point de vue de la veille et de l'évolution dans le temps, les dépendances permettent ici de mettre au jour des évolutions qui affectent les thématiques centrales et leurs évolutions dans un domaine et de cibler les intérêts de recherche dans ce domaine dans une période donnée. Dans ce cas, il ne s'agit ni de repérer précisément des néologismes, ni de repérer des termes « isolés », mais de travailler sur des classes de termes en relation de dépendance qui

permettent de postuler une évolution des connaissances (et plus précisément des centres de recherche et d'intérêt) d'un domaine.

7.2 Extraction des dépendances syntaxiques en corpus

7.2.1 Analyse syntaxique automatique de l'outil Syntex

Dans notre recherche, la notion de dépendance est mise en œuvre à travers l'outil Syntex (Bourigault, 2007 ; Bourigault, *et al.*, 2005) dont on a dit au Chapitre 3 qu'il s'agit d'un analyseur syntaxique dont le fonctionnement est fondé sur l'identification des relations de dépendance entre mots.

Les principales relations de dépendances syntaxiques traitées par Syntex et observées dans cette étude sont présentées dans le Tableau 7.1.

Relation	Tête	Expansion	Exemples
Adjectif	Nom	Adjectif, Participe passé	<i>circuit hybride, circuit imprimé</i>
X-Préposition	Verbe, Adjectif, Nom	Préposition	<i>éclairer en, éclairé sur</i>
Préposition	Préposition	Nom	<i>à semi-conducteur, à batteries</i>
Sujet	Verbe	Nom, Pronom	<i>opérateur appuyer, satellite autoriser</i>
Objet	Verbe	Nom	<i>tourner clé, autoriser satellite</i>

Tableau 7.1 – Principales dépendances syntaxiques observées

Sur la base de ces relations, Syntex identifie des syntagmes et relie en réseaux les termes et syntagmes ainsi identifiés. Bourigault *et al.* (2004 : 14) expliquent :

« Le résultat de l'analyse effectuée par Syntex est un réseau de mots et de syntagmes : un syntagme verbal (resp. nominal, adjectival) est un groupe de mots dont la tête syntaxique est un verbe (resp. nom, adjectif). Par exemple, *révéler une lésion osseuse* est un syntagme verbal dont la tête syntaxique est le verbe *révéler* et l'expansion le syntagme nominal *lésion osseuse*. Ce dernier syntagme a pour tête syntaxique le nom *lésion* et pour expansion l'adjectif *osseuse*. Dans le réseau construit, dit « réseau terminologique », chaque syntagme est relié d'une part à sa tête (lien T) et d'autre part à ses expansions

(lien E) [...]. Les éléments du réseau (mots et syntagmes) sont appelés « candidats termes ». »

Comme nous le verrons dans la suite de ce chapitre, cette notion de réseau est pertinente pour détailler et construire une analyse diachronique. Un extrait de réseau obtenu pour le corpus DORIS est présenté dans la Figure 7.1.

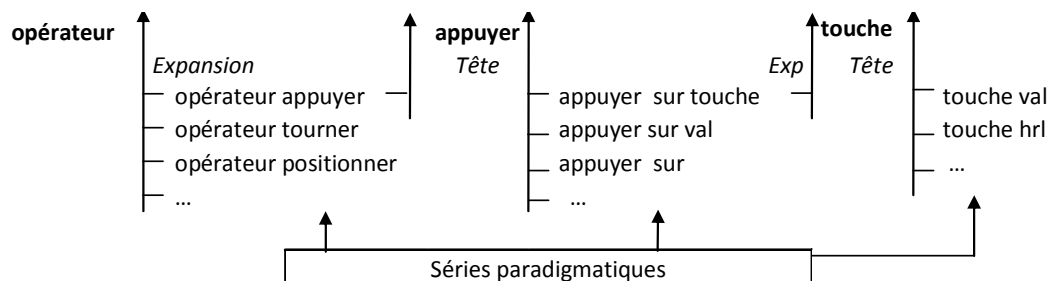


Figure 7.1 – Schéma d'un réseau de dépendances Syntex (extrait DORIS)

Dans la suite de ce chapitre, « série paradigmatique » renvoie à l'ensemble constitué d'un terme et de tous ses descendants. « Terme tête » désigne le terme qui donne naissance à une nouvelle dépendance. Le nouveau syntagme obtenu, en relation de dépendance syntaxique avec le terme tête, est appelé « descendant ».

Les « descendants en tête » et « descendants en expansion » renvoient à deux points de vue sur les dépendances syntaxiques : dans le premier cas, la dépendance est observée à partir d'un terme tête. Par exemple, si *miroir* est un terme tête qui donne naissance à un nouveau descendant tel que *miroir asphérique*, on dit que *miroir asphérique* est un descendant en tête de *miroir*. À l'inverse, *miroir asphérique* est le descendant en expansion de *asphérique*. Ces deux points de vue sur les dépendances sont complémentaires et doivent être pris en compte dans l'analyse.

7.2.2 Sélection des dépendances

Nous l'avons vu, en synchronie, les dépendances sont mises à contribution dans le cadre de la construction de ressources terminologiques. En effet, les dépendances et leur réseau permettent notamment à « l'analyste [de] visualiser des listes paradigmatiques de candidats

termes partageant la même tête ou la même expansion, ce qui le guide vers la constitution de taxinomies locales » (Bourigault, *et al.*, 2004 : 14).

De fait, pour poser un regard diachronique sur les dépendances, comme dans le cas des variantes (Chapitre 6), il est nécessaire de les associer à des empreintes de fréquence spécifiques. Pour cette raison, nous ne sélectionnons dans cette recherche que les cas où un terme tête apparaît dans l'ensemble des sous-corpus mais dont l'ensemble ou une partie des descendants apparaît ou disparaît au fil du temps. Le premier cas de figure est illustré au Tableau 7.2 avec la série paradigmatique de *polissage* et le second au Tableau 7.3, avec la série du terme/concept *laser*.

Comme nous l'avons déjà souligné dans le cas des variantes, le filtre du χ^2 n'est pas pris en compte ici pour sélectionner les fréquences. Celles-ci correspondent simplement au nombre d'occurrences de chaque terme en corpus.

Tête	Descendants	Fréquence (nombre occurrences)		
		TTVS1994	TTVS1998	TTVS2002
<i>polissage</i>		2	28	53
	<i>polissage assisté par ordinateur</i>	0	3	3
	<i>polissage de dioptre</i>	0	2	2
	<i>polissage de lentille</i>	0	2	2

Tableau 7.2 – Exemples de descendants apparus simultanément – *polissage* (TTVS)

Tête	Descendants	Fréquence (nombre occurrences)		
		TTVS1994	TTVS1998	TTVS2002
<i>laser</i>		140	290	342
	<i>laser à cavité étendue</i>	0	8	10
	<i>laser à fibre dopée</i>	0	3	4
	<i>laser à puits quantique</i>	1	1	2
	<i>laser accordable</i>	1	2	2
	<i>laser de communication</i>	1	3	3
	<i>laser monomode</i>	3	5	5
	<i>laser pompé par diode</i>	3	5	3
	<i>laser semi-conducteur monomode</i>	0	2	2
	<i>laser semi-conducteur</i>	8	12	13
	<i>laser solide</i>	8	12	15

Tableau 7.3 – Exemples de quelques descendants apparus – *laser* (TTVS)

Comme l'indique le tableau suivant (Tableau 7.4), les dépendances syntaxiques sont un phénomène important dans nos deux corpus. Ce tableau présente pour chacun des corpus le nombre de termes têtes pour lesquels il existe des dépendances, ainsi que le nombre de termes têtes pour lesquels les descendants entrent dans une empreinte de fréquence pertinente (et que nous sélectionnons donc pour l'analyse).

	Nombre de termes têtes pour lesquels il existe des descendants	Têtes sélectionnées (empreintes de fréquence)
TTVS	418	180
DORIS	183	88

Tableau 7.4 – Sélection des termes têtes qui présentent des descendants (TTVS et DORIS)

Les dépendances syntaxiques concernent 16% des termes sélectionnés dans le TTVS (418 sur les 2557 termes retenus) et plus de 19% des termes de DORIS (183 sur les 937 termes retenus – cf. Chapitre 3). Les empreintes choisies pour cette recherche permettent de sélectionner respectivement 44% des dépendances dans le TTVS et 48% dans DORIS.

En tout, les 180 têtes sélectionnées dans le TTVS produisent 306 descendants et les 88 têtes dans DORIS produisent 174 descendants. En moyenne, sont recensés 2,59 descendants par terme tête dans le corpus TTVS et 2,15 dans DORIS, où chaque tête peut compter de 1 à plus de 10 descendants.

Bien que les dépendances syntaxiques puissent concerner l'ensemble des classes syntaxiques, le nom est naturellement la classe la plus productive et les syntagmes nominaux sont les descendants produits les plus fréquents : 257 têtes sur 418 sont des noms dans le TTVS et 128 sur 183 dans DORIS ; 914 descendants sur 1084 sont des SNom dans le TTVS et 320 sur 394 sont des SNom dans DORIS. La deuxième classe syntaxique la plus productive est le verbe (62 têtes verbales dans le TTVS et 28 dans DORIS).

Le Tableau 7.5 indique pour chaque corpus le nombre de descendants qui apparaissent ou disparaissent dans les sous-corpus les plus récents. Soulignons que les dépendances qui apparaissent/disparaissent simultanément (3^{ème} ligne) renvoient dans le TTVS aux cas où une dépendance disparaît du TTVS1998 mais est présente en 1994 et 2002. Ce type d'observation n'est possible pour chaque descendant que dans la mesure où le corpus TTVS

est composé de 3 sous-corpus. Dans le corpus DORIS, les dépendances ne peuvent qu'apparaître ou disparaître du sous-corpus le plus récent.

	TTVS	DORIS
Descendants qui apparaissent uniquement	278	86
Descendants qui disparaissent uniquement	20	88
Descendants qui apparaissent et disparaissent	8	-
Total	306	174

Tableau 7.5 – Répartition des descendants en fonction des empreintes de fréquence (TTVS et DORIS)

Le tableau suivant (Tableau 7.6) présente la même répartition des dépendances, mais cette fois-ci par rapport au nombre de termes têtes.

	TTVS	DORIS
Têtes dont des descendants apparaissent uniquement	156	34
Têtes dont des descendants disparaissent uniquement	10	37
Têtes dont des descendants apparaissent et disparaissent	14	17
Total	180	88

Tableau 7.6 – Répartition des termes têtes en fonction des empreintes de fréquence de leurs descendants (TTVS et DORIS)

Ces deux tableaux illustrent un aspect déjà mentionné au Chapitre 4 à propos de la croissance de la terminologie dans nos deux corpus. La répartition des descendants apparus et disparus dans ces deux corpus souligne la différence notable entre les deux corpus, que l'on pourrait qualifier de différence entre « renouvellement » et « enrichissement » : dans le TTVS, 91% des cas de descendants apparaissent, contre un peu moins de 50% dans le corpus DORIS. Cette répartition montre que dans DORIS, la terminologie reste équilibrée alors que dans le TTVS, elle est en croissance. Dans le cas du TTVS, on peut donc dire que la terminologie est enrichie alors que dans le corpus DORIS, elle tend plutôt à se renouveler mais à rester stable d'un point de vue quantitatif.

7.3 Observations en corpus

7.3.1 Évaluation de la pertinence globale des dépendances syntaxiques en diachronie

Une première remarque concerne le fait que la pertinence des dépendances syntaxiques pour la diachronie ne doit pas être mesurée en fonction de la pertinence de chacun des descendants isolés, mais doit plutôt prendre en compte les séries paradigmatiques dans lesquelles l'analyse des dépendances apporte des informations sur l'évolution. Ce point est motivé par plusieurs raisons.

La première est qu'analyser chacun des descendants de manière isolée reviendrait à observer l'empreinte de fréquence de chacun comme n'importe quel autre terme. Or, l'intérêt des dépendances syntaxiques est justement de prendre en compte l'analyse d'un terme et du lien de dépendance qu'il entretient avec un autre. Les dépendances apportent des informations centrales pour repérer l'évolution précisément parce qu'elles sont considérées au cœur d'une série paradigmatique.

La seconde raison est méthodologique : nous l'avons vu dans le cas des empreintes de fréquence, il est plus facile pour l'expert et pour l'analyste de porter un jugement sur un groupe de termes que sur des termes isolés. De fait, l'un des intérêts majeurs des dépendances est qu'elles permettent de rassembler plusieurs termes/concepts liés syntaxiquement, ce qui facilite ainsi le jugement sur l'évolution. Ceci est confirmé dans les tableaux suivants où l'on voit que les experts ont pu se prononcer sur l'ensemble des séries paradigmatiques proposées (alors que l'on se souviendra que dans le cas des variantes de termes, plusieurs cas sont restés en suspens).

Pour ces différentes raisons, la pertinence des dépendances est mesurée en fonction de chaque série paradigmatique et non pas pour chaque descendant isolé. C'est ce que présentent les deux tableaux suivants (Tableau 7.7 et Tableau 7.8).

TTVS	Séries sélectionnées	Séries où tous ou certains descendants apparaissent	Séries où tous ou certains descendants disparaissent	Séries où tous ou certains descendants apparaissent/ disparaissent
Pertinentes	93	79	7	7
Non pertinentes	87	77	3	7
Total	180	156	10	14

Tableau 7.7 – Pertinence des séries paradigmatiques (TTVS)

DORIS	Séries sélectionnés	Séries où tous ou certains descendants apparaissent	Séries où tous ou certains descendants disparaissent	Séries où tous ou certains descendants apparaissent/ disparaissent
Pertinentes	27	13	7	7
Non pertinentes	61	21	30	10
Total	88	34	37	17

Tableau 7.8 – Pertinence des séries paradigmatiques (DORIS)

Ces tableaux montrent que la proportion de séries pertinentes dépend du corpus. En effet, si dans le TTVS 51,6% des séries sont jugées pertinentes par les experts, 30,6% le sont pour DORIS.

Bien que tout à fait intéressante, cette proportion de cas pertinents peut paraître un peu décevante par rapport aux trois autres indices. Mais au-delà de ces chiffres, l'intérêt des dépendances réside également dans la diversité des phénomènes qu'elles permettent de mettre au jour, ainsi que dans la richesse des interprétations d'évolution qu'elles permettent de construire. Plus précisément, trois types de phénomènes sont observables à partir des dépendances :

- le repérage de la nouveauté et de l'obsolescence à travers la productivité morphologique et lexicale de certains termes,
- l'apparition de thématiques,
- et le changement de point de vue.

7.3.2 Nouveauté et obsolescence : productivité morphologique et lexicale

Le premier type de phénomène observable relève de ce que Ahmad *et al.* (2002) appellent la productivité morphologique et lexicale (§7.1.2, p.224). Il s'agit de cas où les descendants apparus renvoient à de nouveaux termes/concepts plus spécifiques et dont les dénominations sont créées par composition (« productivité lexicale ») ou par dérivation (« productivité morphologique »). Dans le cas des dépendances analysées par Syntex, les exemples repérés sont principalement ceux de productivité lexicale. Cependant, cet analyseur sait repérer comme dépendances les cas de productivité morphologique où le morphème affixe est marqué par un trait d'union. De fait, bien que la productivité morphologique soit traitée marginalement dans cette analyse, elle n'est pas totalement absente. Le Tableau 7.9 présente un exemple de chacun de ces types de productivité.

Tête	Descendants	Freq TTVS1994	Freq TTVS1998	Freq TTVS2002
<i>circuit</i>		30	42	53
	<i>circuit hybride</i>	0	12	13
<i>satellite</i>		111	206	241
	<i>mini-satellite</i>	0	3	6

Tableau 7.9 – Exemples de productivité morphologique et lexicale (apparition de concepts – TTVS)

Dans ces deux exemples, les termes *circuit hybride* et *mini-satellite* renvoient à des concepts nouveaux dans le domaine spatial. Ici, et en écho à la notion d'interface introduite plus haut (Humbley, 2006), l'intérêt des dépendances est que par rapport aux empreintes de fréquence seules, le lien syntaxique entre la tête et le descendant fournit des informations supplémentaires sur le concept nouveau. Par exemple, dans le cas de *mini-satellite*, l'apport du préfixe indique que la nouveauté réside dans les dimensions de ces satellites, bien moins encombrants que les satellites « classiques ».

Bien que la productivité implique la création d'une unité nouvelle, on peut étendre cette notion aux cas où la productivité s'arrête, c'est-à-dire les cas où une tête « ne produit plus » un ou plusieurs descendants. Ces cas relèvent alors du phénomène d'obsolescence (Tableau 7.10).

Tête	Descendant	Freq DORISGen1	Freq DORISGen3
<i>fonction</i>		20	41
	<i>fonction recalage en temps</i>	40	0

Tableau 7.10 – Exemple de productivité morphologique et lexicale (disparition d’une fonctionnalité – DORIS)

Une autre remarque importante est que si la productivité renvoie essentiellement à un phénomène de création lexicale, dans nos corpus certaines séries paradigmatiques ne rentrent pas dans ce cadre et renvoient à l’apparition (ou disparition) de fonctionnalités ou d’instances, comme nous l’avons décrit dans le Chapitre 4 (§4.3.1.2, p.124*sqq.*). En effet, l’exemple donné ci-dessus (*fonction recalage en temps*) ne signale pas la disparition d’un concept du domaine, mais plutôt celle d’une fonction sur la balise DORIS. On retrouve ici l’influence du domaine d’activité sur l’interprétation, la fonction *recalage en temps* n’étant pas obsolète dans le domaine spatial mais est simplement « enlevée » de l’instrument DORIS. Le Tableau 7.15 présente un autre exemple « d’arrêt » de productivité dans le corpus mais qui cette fois-ci indique la disparition d’une série d’instances (*touche*). Finalement, le Tableau 7.12 présente un exemple d’apparitions de dépendances révélatrices de l’apparition d’une nouvelle fonctionnalité sur l’instrument DORIS (*visibilité satellite*).

Tête	Descendant	Freq DORISGen1	Freq DORISGen3
<i>touche</i>		68	2
	<i>touche eff</i>	6	0
	<i>touche hrl</i>	8	0
	<i>touche rct</i>	4	0
	<i>touche seq</i>	6	0
	<i>touche syn</i>	6	0
	<i>touche tst</i>	7	0
	<i>touche val</i>	22	0

Tableau 7.11 – Exemple de productivité morphologique et lexicale (disparition d’instances – *touche*) (DORIS)

Expansion	Descendants	Fréquence	
		DORISGen1	DORISGen3
<i>satellite</i>		11	131
	<i>activation satellite</i>	0	9
	<i>altitude de satellite</i>	0	3
	<i>autorisation satellite</i>	0	10
	<i>coordonnée satellite</i>	0	3
	<i>identification de satellite</i>	0	3
	<i>orbite de satellite</i>	0	5
	<i>passage de satellite</i>	0	4
	<i>vecteur satellite</i>	0	4
	<i>visibilité de satellite</i>	0	4
	<i>visibilité satellite</i>	0	8

Tableau 7.12 – Exemple de productivité morphologique et lexicale (apparition de fonctionnalité – *visibilité satellite*) (DORIS)

L'exemple de la série paradigmatique de *satellite*¹⁰⁴ permet de bien montrer la « relativité » de la nouveauté : dans le corpus DORIS, bon nombre de termes/concepts apparus relèvent en effet non pas de la création de termes/concepts dans un domaine de connaissance, mais de la création de termes/concepts dans un domaine d'activité. Ces termes/concepts ne sont donc nouveaux que du point de vue du projet DORIS, et non pas dans le domaine spatial en général.

Un autre intérêt de ces exemples est que l'apparition/disparition de leurs descendants se manifeste « massivement » puisque tout un ensemble de descendants apparaît/disparaît simultanément. Néanmoins, lorsque ce type de phénomène ne peut être mis en lien avec la notion de nouveauté ou d'obsolescence (de concepts, d'instances ou de fonctionnalités), l'apparition massive de dépendances doit être mise en lien avec la centralité ou au contraire le retrait ponctuel de thématiques.

7.3.3 Apparition/disparition de thématiques

Ce phénomène correspond aux cas où plusieurs descendants apparaissent/disparaissent simultanément sans que l'on puisse les mettre en lien avec les notions de nouveauté ou

¹⁰⁴ Cf. Chapitre 4 (§4.3.1.2, p.124) pour une description détaillée des causes d'apparition de cette fonctionnalité.

d'obsolescence. Ce que l'on repère plutôt dans ce cas précis est l'apparition/disparition de thématiques. Ce phénomène, déjà décrit dans le cas des variantes (Chapitre 6) et des empreintes de fréquence (Chapitre 4), renvoie à la centralité ponctuelle de thématiques connues ou au retrait de thématiques qui ne sont pourtant pas obsolètes.

Cet aspect est particulièrement remarquable dans le corpus TTVS, dans lequel il est aisé de suivre le développement de nouveaux projets spatiaux qui entraînent la mise en avant de questions sur les développements de ces projets. On retrouve ici une thématique déjà rencontrée dans notre étude, celle des miroirs, dont une série de descendants du terme tête *miroir* apparaissent simultanément à partir de 1998.

Tête	Descendants	Fréquence		
		TTVS1994	TTVS1998	TTVS2002
<i>miroir</i>		88	205	295
	<i>miroir concave correcteur arrière</i>	0	2	2
	<i>miroir concave correcteur</i>	0	3	3
	<i>miroir hors d'axe</i>	0	1	8
	<i>miroir mère</i>	0	0	8
	<i>miroir secondaire</i>	0	3	4
	<i>miroir sélectif</i>	0	2	2
	<i>miroir sphérique</i>	0	3	3

Tableau 7.13 – Apparition d'une série paradigmatique – thématique des miroirs (TTVS)

L'exemple de la série paradigmatique *miroir* reflète les nombreux progrès qui ont vu le jour sur les questions de l'élaboration et du traitement des miroirs : en effet, aucun de ces descendants ne peut être mis en lien avec la notion de nouveauté. Ce sont des termes/concepts connus qui sont rappelés dans le corpus pour traiter de la question des miroirs, centrale dans certains projets développés dans le TTVS. Du point de vue de la veille scientifique et technique, ce type de série est intéressant dans la mesure où, bien qu'aucun de ces termes/concepts ne soient nouveaux, la série indique l'intérêt du domaine pour le concept de *miroir*. La question est de plus en plus développée, ce qui se traduit par le rappel en corpus de termes liés à ce concept. C'est ce que nous avons appelé dans le chapitre précédent *centralité*.

Un autre exemple est présenté ci-dessous :

Tête	Descendants	Fréquence		
		TTVS1994	TTVS1998	TTVS2002
<i>sondeur</i>		5	8	10
	<i>sondeur passif</i>	0	2	2
	<i>sondeur vertical</i>	0	3	3
	<i>sondeur au limbe</i>	0	3	4

Tableau 7.14 – Apparition d’une série paradigmatique – thématique des sondeurs (TTVS)

La série paradigmatique obtenue à partir du terme tête *sondeur* permet d’illustrer un autre cas d’activité en croissance depuis quelques années, lié cette fois-ci à de nouveaux projets et missions spatiales clairement identifiables au Cnes tels que Pharaon ou GOMOS. Ces descendants indiquent la centralité de la question des sondeurs à partir de 1998 au Cnes, en lien avec ces nouveaux projets qui entraînent l’apparition de ces thématiques dans le cours pour développer ces questions dans le domaine.

L’autre élément remarquable en lien avec l’apparition de ces séries paradigmatiques est qu’il arrive que celles-ci ne soient pas isolées mais soient au contraire accompagnées par d’autres séries paradigmatiques thématiquement liées. Si l’on reprend l’exemple de la thématique des miroirs (Tableau 7.13, p.241), on peut observer que les descendants obtenus donnent à leur tour naissance à d’autres descendants qui apparaissent massivement. En effet, dans cet exemple, le descendant *miroir concave correcteur* permet de mettre au jour au moins trois autres séries paradigmatiques qui apparaissent à partir de 1998 :

- *miroir concave correcteur* peut être lié à la série paradigmatique de *miroir concave* dans la mesure où il est un de ses descendants en tête ; trois autres descendants apparaissent dans cette série : *décentrement de miroir concave*, *décentrement miroir concave* et *déplacer miroir concave*,
- ce terme/concept peut également être lié à la série paradigmatique de *correcteur* dont il est un des descendants en expansion et qui compte deux descendants apparus : *lentille de correcteur* et *miroir concave correcteur*,
- *miroir* peut enfin être relié à la série paradigmatique de *polissage* qui contient 4 descendants apparus : *polissage assisté par ordinateur*, *polissage de dioptre*, *polissage de miroir*, *polissage de lentille*.

Dans le cas de la centralité de thématiques, plusieurs séries paradigmatiques peuvent donc indiquer la centralité d'une thématique donnée. L'analyse des dépendances peut donc permettre de tisser des liens entre plusieurs séries pertinentes pour renforcer l'interprétation de la centralité. Ces liens permettent également d'affiner l'interprétation. Par exemple, toujours dans le cas de la thématique des miroirs, le lien entre la série paradigmatique de *miroir* et de *polissage* permet de déduire que les techniques de polissage ont évolué et que cette question participe à la centralité de la question des miroirs à partir de 1998. Cet exemple illustre l'importance du réseau de dépendances tels que représenté à la Figure 7.1 (p.232) et dont on peut tirer parti pour repérer l'évolution en corpus. Il permet également d'insister sur le fait que les dépendances « en tête » et « en expansion » sont deux points de vue complémentaires sur le corpus que l'on doit combiner et prendre en compte pour analyser l'ensemble des informations fournies *via* les dépendances syntaxiques.

Enfin, le fait que l'évolution ait tendance à se tisser « en réseau » est une caractéristique dont on a pu sentir les effets pour chacun des quatre indices, mais qui est particulièrement marquée et pertinente dans le cas des dépendances. Il est donc nécessaire de chercher à en tirer profit d'un point de vue méthodologique, comme nous le verrons au Chapitre 9.

7.3.4 Changement de point de vue

Le troisième et dernier phénomène lié aux dépendances syntaxiques observable dans nos corpus permet d'aborder la question du changement sémantique. Plus précisément, nous parlerons de changement de point de vue.

Nous empruntons cette notion des travaux de Condamines et Rebeyrolle (1997a ; Rebeyrolle, 1996) qui ont exploré dans le domaine spatial la manière dont, en discours, l'interlocuteur « peut se rendre compte de la signification pour laquelle a opté le locuteur » (Condamines & Rebeyrolle, 1997a) lorsque le terme utilisé est polysémique. Les auteures détaillent l'exemple du terme *satellite* dans des textes de deux divisions différentes au Cnes, la division « Observation de la Terre » et « Mathématiques spatiales ». Grâce à la mise au jour de six patrons lexico-syntaxiques différents, elles identifient six acceptions différentes pour le terme *satellite* (*ibidem* : 181-182) :

1. le satellite peut être considéré comme un corps artificiel,
2. le satellite est considéré comme un objet mobile dont on peut changer la position,
3. le satellite peut être considéré comme une plateforme (sachant que la plateforme est également une partie du satellite et que l'on se trouve dans un cas de glissement de sens par métonymie),
4. encore une fois par métonymie, le satellite peut être considéré comme un véhicule,
5. le satellite peut être considéré comme un hôte,
6. le satellite peut être considéré comme un relais.

Ces six points de vue sur le terme/concept *satellite* sont identifiables à partir des structures lexico-syntaxiques dans lesquelles il est construit en discours, et ne reflètent pas réellement un fonctionnement polysémique, mais plutôt différents points de vue sur cet objet :

« Ce que nous retiendrons des [six] types de constructions mis en évidence c'est que le terme *satellite* n'est pas véritablement polysémique mais plutôt que coexistent différentes façons de l'aborder, différents angles depuis lesquels on peut considérer cet objet [...]. Il serait [...] erroné de dire que le terme *satellite* est monosémique, néanmoins, bien qu'il ne s'agisse pas non plus de polysémie, le mot reste aussi ambigu que dans la langue générale. »
(Rebeyrolle, 1996 : 49)

La notion de *point de vue* proposée par les auteures est intéressante à double titre. Le premier argument est que cette notion s'inscrit dans une perspective discursive : les différents points de vue sur un terme donné sont observables en discours, à travers notamment la combinatoire des termes dans les textes produits. Cet aspect implique aussi que les points de vue sont associables à différents groupes de locuteurs identifiables, et que les discours varient en fonction de ces locuteurs, de leurs relations et des situations de communication (Beacco & Moirand, 1995 : 34). Transposé sur la dimension diachronique, cela signifie que les discours sont adaptés dans le temps, en fonction des locuteurs impliqués dans le discours, de la période à laquelle ils sont produits, et surtout des connaissances dont les locuteurs disposent sur les termes/concepts impliqués dans ces discours.

Le second élément est que cette notion permet de construire une représentation de la polysémie en tant que phénomène continu. Plus exactement, et comme la citation de Rebeyrolle ci-dessus l'indique, il est possible de mettre au jour des fonctionnements qui ne relèvent ni de la monosémie, ni de la polysémie et que l'on peut définir comme des points de vue différents pouvant coexister sur le même objet. L'intérêt de cette vision sur le changement de sens est d'autant plus pertinent en diachronie courte que le regard que l'on porte sur l'évolution implique un grain très fin d'analyse, peut-être trop fin pour observer l'émergence complète de nouveaux sens établis.

C'est ce que nous illustrons à travers deux exemples d'analyses dans les corpus DORIS et TTVS. Un premier cas peut être illustré par l'exemple d'*opérateur* dans le corpus DORIS, terme pour lequel on observe un changement de dépendances marqué (Tableau 7.15).

Expansion	Descendants	Fréquence	
		DORISGen1	DORISGen3
<i>opérateur</i>		89	26
	<i>interface opérateur</i>	0	12
	<i>opérateur appuyer sur touche</i>	26	0
	<i>opérateur appuyer sur val</i>	5	0
	<i>opérateur appuyer</i>	33	0
	<i>opérateur tourner clé</i>	6	0

Tableau 7.15 – Exemple de changement de point de vue – le cas d'*opérateur* (DORIS)

Dans cet exemple, une première remarque concerne le fait que le nouveau descendant *interface opérateur* renvoie à un élément nouveau sur la balise DORIS. Mais au-delà de cet aspect, il est intéressant d'observer la disparition massive de toutes les dépendances verbales où le verbe implique une action manuelle dont l'opérateur est l'agent. Cet élément laisse entrevoir le changement du rôle de l'opérateur sur la balise DORIS : il disparaît en tant qu'agent d'une action sur la balise (*appuyer, tourner la clé*) mais apparaît associé au terme *interface*. Le néologisme *interface-opérateur* révèle l'influence de l'informatique dans le fonctionnement des balises et sur le rôle de l'opérateur. Ce dernier a grandement évolué en

10 ans et l'on peut aisément émettre l'hypothèse que la formation et les responsabilités de l'opérateur DORIS ne sont plus les mêmes aujourd'hui¹⁰⁵.

Bien que le changement sémantique associé au terme/concept *opérateur* soit sensible, on ne peut parler de polysémie. L'opérateur reste l'agent humain qui intervient sur la balise, mais son rôle et ses « moyens d'intervention » sont différents en 2000 (génération 3) par rapport à 1989 (génération 1).

Un second exemple est tiré du TTVS. Il s'agit du cas de l'adjectif *parasite*. Dans ce cas, à l'inverse d'*opérateur*, la série paradigmatique du terme n'est pas renouvelée, mais précisée et complétée par l'apparition de nouveaux descendants (Tableau 7.16).

Expansion	Descendants	Fréquence		
		TTVS1994	TTVS1998	TTVS2002
<i>parasite</i>		40	87	142
	<i>lumière parasite</i>	18	34	66
	<i>flux parasite</i>	3	16	8
	<i>source parasite</i>	3	5	3
	<i>élément parasite</i>	2	5	6
	<i>signal parasite</i>	0	3	5
	<i>rayonnement parasite</i>	0	0	4
	<i>image parasite</i>	0	7	17
	<i>flux lumineux parasite</i>	0	4	2
	<i>trajet parasite</i>	0	2	5

Tableau 7.16 – Exemple de changement de point de vue – le cas de *parasite* (TTVS)

Ce tableau illustre l'apparition de nouvelles dépendances contenant l'expansion *parasite*, qui sont plus spécifiques que les dépendances déjà existantes en 1994. En effet, *flux*, *source*, *lumière* et *éléments parasites* sont complétés par de nouveaux éléments parasites tels que *image*, *trajet* ou *signal parasites* qui n'apparaissent qu'en 2002. Il semble alors qu'avec le temps, les locuteurs ont recours à des termes/concepts beaucoup plus diversifiés pour appréhender la notion de *parasite*.

Si l'on complète l'analyse des dépendances Syntex par une observation en corpus, on s'aperçoit rapidement que les verbes dont l'un des arguments est un SNom qui

¹⁰⁵ L'analyse de cet exemple sera complétée dans le Chapitre 9 (§9.1.3, p.281).

contient l'expansion *parasite* sont eux aussi grandement diversifiés et précisés. En effet, en 1994, les contextes verbaux contenant *parasite* sont rares et se limitent à décrire le fait que les éléments parasites sont indésirables, qu'il faut les extraire ou les éviter¹⁰⁶ :

- (lxxv) ... de **rejeter** le maximum de lumière parasite, ...
 ... chaque détecteur **ne doit pas recevoir** de flux parasite...
 ... pour **extraire** des éléments parasites...

En 2002, ces contextes sont beaucoup plus nombreux et variés et montrent des points de vue nouveaux sur cette question. En effet, bien qu'il soit toujours nécessaire d'extraire ou de supprimer ces parasites (exemples (lxxvi)), on entrevoit le développement de nouvelles solutions marquées par l'emploi de verbes tels que *permettre* ou *arriver à* (exemples (lxxvii)) qui sont utilisés pour expliquer les nouvelles solutions existantes.

- (lxxvi) ... les logiciels permettent **d'extraire** l'ensemble des éléments parasites...
 ... permet de **diminuer** les éléments parasites ce qui permet d'améliorer les...
 ... élémentaire ainsi que la **réduction** des éléments parasites permettent d'accroître...
 (lxxvii) ... permet de **s'affranchir** des parasites. Des essais sur l'instrument complet permettent...
 ... reflets permettent également **d'éliminer** les images parasites qui peuvent se...
 ... permet de **réduire** les réflexions parasites par une géométrie soignée...
 ... utile, de façon également à **limiter** le flux parasite. Le refroidissement...
 ... pour **arriver à** une bonne **suppression** de cette lumière parasite...

Mais surtout, ces nouveaux contextes indiquent qu'il ne s'agit plus seulement d'extraire ou de supprimer les parasites, mais de les *évaluer*, de les *identifier* ou de les *maîtriser* (exemples (lxxviii)) :

- (lxxviii) ... pour **évaluer** les parasites engendrés par l'image et les **soustraire**...
 ... sur la puce est figée certains éléments parasites **sont mal maîtrisés**...
 ... On **identifie** alors les images parasites par un tracé de rayons non séquentiel...

¹⁰⁶ Soulignons que ce type d'observation relève toujours du champ des dépendances syntaxiques. En effet, ce que nous observons à travers ces contextes sont des modifications des dépendances de chacun des verbes soulignés en gras dans les extraits. Cependant, chacun de ces contextes est trop rare dans le TTVS pour que Syntex repère ces dépendances automatiquement. La petite taille du corpus révèle dans ce cas une limite certaine pour observer des évolutions sémantiques, et l'on doit compléter l'analyse par une analyse manuelle en corpus.

D'autres contextes permettent d'affiner l'explication de cette évolution dans la mesure où ils mettent en évidence l'existence de nouvelles possibilités liées à de nouveaux moyens de modélisation informatisée de ces parasites (exemples (lxxix)) :

(lxxix) ... ***Modéliser** la lumière parasite dans un instrument requiert des logiciels...*
 ... On ***identifiera*** les trajets parasites par modélisation...

Il est donc clair qu'en 2002, il ne s'agit plus d'avoir conscience de ce phénomène indésirable, mais d'apporter des solutions fines et efficaces pour lutter contre eux et les éliminer.

Enfin, si l'on analyse le terme *lumière parasite* qui est le syntagme le plus fréquemment formé avec l'adjectif *parasite*, on observe le même type de changements contextuels, résumés dans le Tableau 7.17.

TTVS1994	TTVS2002
filtrage de ~ piégeage de ~ suppression de ~	filtrage de ~ piégeage de ~ suppression de ~ lutter contre ~ réduire ~ contrôle de ~ spécification de ~ étude de ~

Tableau 7.17 – Changement de point de vue – *lumière parasite* (TTVS)

Dans cet exemple encore une fois, il serait trop hâtif de parler de polysémie. Néanmoins, il est clair que les connaissances des experts sur le terme/concept *parasite* ont augmenté et que lorsque cette question est abordée en 2002, elle l'est de manière plus complète et précise qu'en 1994. En d'autres termes, lorsque le locuteur aborde la question des parasites en 2002, il le fait en tenant compte des connaissances actuelles des interlocuteurs sur cette question, de leur formation, etc., qui sont différentes de celles de 1994. Ceci illustre bien le fait que le point de vue est intimement lié à la notion de discours et aux acteurs qui y prennent part.

Une différence, sur laquelle nous revenons dans le chapitre suivant, existe cependant entre les cas de *parasite* et d'*opérateur* : dans le premier cas, le concept *parasite* n'est pas modifié. Les experts découvrent progressivement les propriétés et la nature de ces parasites, mais le concept en lui-même n'évolue pas. À l'inverse, on peut dire que le concept *opérateur* est modifié, ne serait-ce qu'au niveau des tâches que cet agent accomplit sur la balise.

Si cette distinction ne remet pas en cause les phénomènes de changements de points de vue décrits ici, il peut être important de la prendre en compte pour décider si les changements observés impliquent l'apparition d'un nouveau sens ou non. En effet, puisque la notion de *point de vue* permet d'adopter une vision continue sur le changement sémantique dans le temps, il est clair qu'à un point donné sur ce continuum entre monosémie et polysémie on peut basculer du changement de point de vue vers l'émergence d'un nouveau sens dans la langue. Identifier, ou peut-être plutôt décider, l'emplacement de ce point n'est pas chose aisée et relève très certainement encore une fois de l'interprétation de l'analyste et des experts. En effet, si l'émergence d'un nouveau sens ou d'un nouveau point de vue peut être identifiée à travers la mise au jour de régularités lexico-syntaxiques en discours, il reste à la discrétion des interprètes et des analystes de ce discours de décider si ces régularités manifestent l'apparition d'un nouveau sens ou de l'existence de différents points de vue sur le même objet.

7.4 Bilan

Ce dernier indice proposé pour repérer l'évolution des connaissances d'un domaine en corpus se révèle une fois encore très riche en informations.

D'un point de vue méthodologique, la notion de dépendances syntaxiques est un indice très prometteur mais qu'il faut manipuler avec précaution.

Tout d'abord, les dépendances relèvent de différentes classes syntaxiques, dont on peut émettre l'hypothèse qu'elles ne reflètent pas les mêmes phénomènes linguistiques. Par exemple, nous avons traité l'ensemble des dépendances sans les distinguer *a priori*, mais il est clair que les SV et les SN n'ont pas le même statut. À propos des SV, pour reprendre une

remarque de L'Homme (1998), les verbes revêtent un double intérêt en terminologie dans la mesure où ils peuvent soit être des termes du domaine (par exemple *détecter* dans le TTVS), soit apporter des informations sur la distribution des termes (par exemple *appuyer* dans DORIS). Dans les cas présentés dans ce chapitre, les SV analysés relèvent tous plutôt de cette seconde catégorie. Dans ce cas, il n'est pas pertinent d'observer ces SV pour repérer des cas de productivité morphologique et lexicale par exemple. De fait, bien qu'il soit important de conserver l'ensemble des classes syntaxiques pour observer les dépendances, on pourra chercher à affiner leur impact en fonction de ces classes. Mais déjà, ces exemples de dépendances permettent de montrer l'importance des classes syntaxiques non nominales qui, bien que souvent négligées en terminologie, se révèlent centrales pour l'analyse.

Deuxièmement, l'intérêt et la richesse des dépendances syntaxiques pour la description linguistique entraînent un certain flou sur la manière dont elles peuvent être mises en œuvre. Nous l'avons illustré au §7.1.2 (p.224), le recours aux dépendances peut être envisagé dans une optique strictement harrissienne, dans une simplification des principes harrisiens pour l'extraction automatique de classes sémantiques ou encore pour l'observation de la distribution de termes spécifiques. Si chacune de ces démarches revêt un intérêt en diachronie, toutes ne peuvent être facilement mises en œuvre et ne requièrent pas le même type de corpus (en particulier la taille). Dans notre recherche, nous avons centré nos analyses sur l'observation de termes isolés et de séries paradigmatiques et non pas sur l'identification de classes sémantiques. Il s'agit d'une des démarches les plus simples à reproduire en corpus, qui se démarque par la richesse des éléments d'analyse qu'elle fait émerger.

Troisièmement, il s'agit d'un indice particulièrement fructueux lorsque l'on travaille avec les experts. En effet, les séries paradigmatiques permettent de regrouper plusieurs termes en « familles » (lexicales, morphologiques mais aussi conceptuelles), ce qui aide à la fois l'analyste et les experts à proposer une interprétation. Les séries paradigmatiques permettent en effet de porter un jugement sur des ensembles de termes liés, ce qui « rassure » l'expert et lui donne des arguments supplémentaires pour justifier ses choix. Et il

est clair, en effet, que les dépendances sont l'un des indices sur lesquels la collaboration avec les experts a été la plus aisée.

Enfin, terminons par le fait que les dépendances se révèlent particulièrement riches d'un point de vue diachronique et permettent de mettre au jour des phénomènes d'évolution d'au moins trois ordres : la néologie, la centralité/le retrait de thématiques et les changements de point de vue. Il n'est pas toujours aisé de distinguer ces phénomènes (en particulier les deux premiers) et une des perspectives est de pouvoir combiner les dépendances à d'autres indices pour pouvoir trancher entre ces différents cas. En effet, si l'on peut différencier certains changements de points de vue en observant la nature des descendants en jeu (les SV principalement), on peut imaginer que d'autres indices, tels que la présence de contextes riches en connaissances d'évolution, peuvent permettre de mieux interpréter les dépendances en diachronie. Cette piste de combinaison d'indices, déjà mentionnée plus tôt dans notre étude, est centrale et fait l'objet d'un chapitre dans la partie suivante. Néanmoins, même si l'observation de l'évolution se concentre sur ce seul indice, il est clair que celui-ci permet de mettre en place des analyses fines d'évolution, y compris sur les changements sémantiques, qui sont parmi les plus difficiles à mettre au jour.

De manière générale, la description de ces quatre indices permet de montrer les différentes facettes que revêt l'évolution des connaissances en diachronie courte. Loin d'être un phénomène négligeable, l'évolution sur de courtes périodes se révèle particulièrement riche, bien qu'elle intervienne à un grain très fin. Dans le chapitre suivant, nous proposons un bilan de ces multiples facettes et proposons une typologie de l'évolution des connaissances en diachronie courte.

Chapitre 8 Typologie de l'évolution en diachronie courte

SOMMAIRE DU CHAPITRE

8.1	TYPLOGIE DE L'ÉVOLUTION EN DIACHRONIE COURTE.....	254
8.2	RETOUR SUR LES DIMENSIONS INTERNE ET EXTERNE	261
8.2.1	<i>Dénomination, concept, instance.....</i>	<i>261</i>
8.2.2	<i>Différents niveaux d'évolution.....</i>	<i>262</i>
8.2.3	<i>Évolution des techniques et évolution des sciences.....</i>	<i>264</i>
8.2.4	<i>Évolution et continuité.....</i>	<i>267</i>
8.2.4.1	Entre discours et langue	267
8.2.4.2	Du même à l'autre	268
8.3	CONSÉQUENCES MÉTHODOLOGIQUES.....	270

Les descriptions proposées dans les chapitres précédents ont permis de montrer la pertinence de l'hypothèse d'un parallèle entre langue et connaissances pour repérer l'évolution. À la lueur des analyses menées, il est non seulement clair que les quatre indices linguistiques étudiés permettent de repérer des évolutions de connaissances pertinentes, mais également que ces évolutions sont multiples et variées. Ceci vient confirmer le fait que l'évolution des connaissances spécialisées est un phénomène central et très rapide observable en diachronie courte, mais aussi qu'il présente de multiples facettes. Pour synthétiser ces observations, ce chapitre dresse un bilan des différents types d'évolution que nous avons pu observer dans nos deux corpus et revient sur les principaux aspects du parallèle interne/externe à prendre en compte pour l'interprétation.

Dans la première section, nous proposons une typologie des différentes évolutions rencontrées en diachronie courte dans le domaine spatial. Dans la deuxième section, nous faisons un point sur la question des dimensions interne et externe pour la diachronie et sur les choix qui s'imposent à l'analyste pour mener la description. En guise de transition avec la dernière partie de ce manuscrit, nous terminons par les conséquences méthodologiques impliquées par ces différents aspects.

8.1 Typologie de l'évolution en diachronie courte

L'évolution en diachronie courte est un point de vue inédit en langue de spécialité et l'un de nos objectifs était précisément de caractériser l'évolution en jeu dans ce contexte. La description des quatre indices nous a permis de mettre au jour différents aspects de l'évolution des connaissances en diachronie courte. Nous proposons dans cette section un bilan des 17 interprétations possibles dégagées à partir de l'analyse des corpus DORIS et TTVS. Chaque type d'évolution est illustré d'un ou plusieurs exemple(s) rencontré(s) dans notre étude et associé(s) aux indices ayant permis de mettre au jour cette interprétation.

Ces 17 catégories sont résumées sous forme de tableau en Annexe C (p.386).

1. Apparition d'un nouveau concept dans le domaine

Ce groupe rassemble les cas classiques de néologie où un terme/concept nouveau apparaît dans le domaine. Ne sont repérés dans nos corpus que des cas où le terme (dénomination) et le concept apparaissent simultanément, sachant qu'un concept seul peut apparaître et se « greffer » à une dénomination existante¹⁰⁷. On peut parler dans ce cas de néologie de sens. Rappelons que, bien que cette évolution soit souvent mise en avant en terminologie, elle est peu fréquente dans nos corpus.

Nous rapprochons cette interprétation de la notion de *récence*, parfois proposée par nos experts. La nuance est ténue, mais met en avant la continuité sur laquelle se construit l'évolution. En choisissant le terme de *récence*, les experts marquent une certaine familiarité avec le terme/concept observé, bien que celui-ci puisse éventuellement être considéré comme nouveau.

Exemples : *APS, mini-satellite* (TTVS)

Indices pertinents : Chaque indice observé est susceptible de nous mener à ce type d'évolution et d'indiquer la présence d'un néologisme : la fréquence (apparition d'un terme isolé), la présence de contextes

¹⁰⁷ Cependant, ce type de phénomène n'a pu être observé dans nos corpus. On peut émettre l'hypothèse que les intervalles de temps observés sont trop petits. Cela devra être confirmé par de nouvelles analyses sur de nouveaux corpus.

riches en connaissances évolutives, les variantes et les dépendances.

2. Obsolescence d'un concept dans le domaine

Il s'agit du pendant de la néologie qui regroupe les cas de disparition de termes/concepts.

Exemples : *LDEF (projet TTVS), détecteur pneumatique (TTVS)*

Indices pertinents : Nous n'avons que très peu d'exemples de termes/concepts obsolètes dans nos corpus, mais dans ces cas, ce sont l'indice de fréquence¹⁰⁸ (disparition d'un terme isolé), la présence de contextes riches en connaissances évolutives et certaines dépendances qui permettent de les repérer.

3. Apparition non planifiée d'une variante

Cette catégorie rentre dans le cadre de la néologie de forme. Il s'agit de cas où la dénomination d'un concept change dans le temps, sans que cette modification soit planifiée et sans que cela affecte le concept. Typiquement, on retrouve de nombreux cas de réduction de termes et de siglaison.

Exemples : *coffret de batterie → coffret batterie (DORIS), rapport signal-sur-bruit → S/B (TTVS)*

Indices pertinents : Les deux indices qui permettent de repérer ces cas dans nos corpus sont l'indice de fréquence (apparition d'un terme isolé) et les variantes.

4. Disparition non planifiée d'une variante

Ces cas se rapprochent d'un phénomène d'implantation : alors que plusieurs dénominations existent pour un concept donné, certaines se stabilisent au détriment d'autres qui

¹⁰⁸ Dans le cas de l'indice de fréquence rappelons que la fréquence observée n'était pas statistiquement significative (calcul du χ^2 , p.111) et l'exemple que nous proposons ici a été discuté entre les experts.

disparaissent, sans que le changement soit planifié et sans que cela affecte le concept mais sans non plus qu'aucune politique linguistique n'ait été envisagée pour cela.

Exemples : *synthèse d'ouverture optique* (→ *synthèse d'ouverture*) (TTVS)

Indices pertinents : Les deux indices qui permettent de repérer ces cas dans nos corpus sont l'indice de fréquence (disparition d'un terme isolé) et les variantes.

5. Apparition planifiée d'une variante

Il s'agit de cas, relativement rares au Cnes, où une commission décide de modifier la dénomination d'un concept de manière planifiée. Ce cas peut également être rapproché des cas de néologie de forme.

Exemples : *anomalie* (DORIS)

Indices pertinents : Cf. Apparition non planifiée d'une variante.

6. Disparition planifiée d'une variante

Dans la mesure où de nouvelles dénominations sont proposées pour un concept, les anciennes disparaissent (de manière planifiée).

Exemples : *panne* (DORIS)

Indices pertinents : Cf. Disparition non planifiée d'une variante.

7. Apparition d'une instance

Cette catégorie regroupe les cas où l'élément nouveau est une instance de concept. Nous n'avons pas repéré d'exemple attesté en corpus, mais le répertorions comme cas potentiel.

8. Disparition d'une instance

Dans ce cas, une instance spécifique disparaît du domaine (essentiellement un domaine d'activité), mais le terme/concept n'évolue pas dans le domaine.

Exemples : *interrupteur secteur, touche VAL (DORIS), Modèle Instrumental (TTVS)*

Indices pertinents : L'indice qui permet de repérer ces cas est essentiellement l'indice de fréquence (la disparition d'un terme isolé et, dans le cas de DORIS, le regroupement de formes chrono-homogènes disparues), complété parfois par les dépendances syntaxiques.

9. Apparition de fonctionnalités

Cette catégorie regroupe les cas où des fonctionnalités sont nouvelles pour un projet ou pour concept donné, sans être nouvelles dans le domaine en général. Ce type d'interprétation est caractéristique des domaines d'activité et des corpus projet dont on a dit que les connaissances représentées ne concernent pas l'ensemble d'un domaine de connaissance, mais un sous-ensemble propre à un contexte scientifique donné (souvent lié à une application, Chapitre 3, p.79). Ces cas se distinguent donc des néologismes classiques en ce qu'ils renvoient à des nouveautés uniquement relatives à un projet ou une application donnée, mais qui sont connues dans le domaine en général. Nous avons choisi de parler de « nouvelles fonctionnalités » dans la mesure où les cas relevés dans nos corpus renvoient tous à des fonctionnalités. Néanmoins, on peut dire qu'il s'agit d'une néologie relative à un contexte appliqué donné et on pourrait par exemple les qualifier de « néologismes relatifs ». L'intérêt de cette classe est surtout de marquer l'importance de la relativité de la nouveauté en fonction des connaissances représentées dans les corpus analysés, ainsi que de bien distinguer ce qui relève de la néologie dans un domaine de connaissance et ce qui relève de la nouveauté dans un contexte précis.

Exemples : *visibilité satellite (DORIS)*

Indices pertinents : Les indices qui permettent de repérer ces cas sont l'indice de fréquence (formes chrono-homogènes et rupture de formes isolées), les contextes riches en connaissances évolutives et la productivité morphologique et lexicale.

10. Disparition de fonctionnalités

Sur le modèle de la catégorie précédente, cette catégorie regroupe les fonctionnalités qui disparaissent d'un domaine d'activité.

Exemples : *fonction recalage en temps (DORIS)*

Indices pertinents : Cf. Apparition de fonctionnalités.

11. Stabilisation/implantation d'un concept (familiarisation)

Cette catégorie regroupe les cas de termes/concepts dont on observe la finalisation de l'implantation dans le domaine. Il s'agit de termes/concepts dont on peut dire qu'ils sont parfaitement intégrés dans le domaine et sont arrivés à une certaine maturité. Ces cas ne sont observables qu'à partir du moment où l'on a pu mettre au jour des traces de l'instabilité de ces termes/concepts à une période antérieure dans le domaine.

Exemples : *Bi-CMOS (→ BiCMOS) (TTVS), oscillateur ultra-stable (→ OUS) (DORIS)*

Indices pertinents : Les indices qui permettent d'observer ces cas sont les variantes, associées à la fréquence pour la mise au jour de schémas d'implantation.

12. Migration de termes/concepts d'un domaine connexe

Nous l'avons vu à travers l'exemple de l'informatique dans le projet DORIS, les frontières entre domaines sont perméables. En diachronie, ceci se traduit entre autres par la migration de termes/concepts. D'un point de vue terminologique, ceci signifie que le stock terminologique d'un domaine se voit enrichi, mais du point de vue extralinguistique, cette migration est susceptible d'entraîner des modifications importantes dans le domaine qui « reçoit » ou « emprunte » ces termes/concepts (par exemple, dans DORIS, nous avons vu que les tâches de l'opérateur sur la balise avaient changé et bénéficié de l'informatisation).

Exemples : Termes en lien avec le domaine informatique : *fichier, téléchargement, interface homme-machine, IHM, etc. (DORIS)*

Indices pertinents : Les indices qui permettent d'observer ces cas sont essentiellement la fréquence (termes isolés et termes chrono-homogènes) et les dépendances.

13. Centralité ponctuelle de thématiques anciennes

Nous l'avons vu en particulier dans le cas du TTVS, il arrive que des termes/concepts liés à une thématique apparaissent simultanément dans le corpus. Il ne s'agit ni de termes/concepts nouveaux ni de thématiques nouvelles, mais ils ont dû être rappelés pour aborder d'autres progrès, nouveaux ou récents cette fois-ci. Ceci est particulièrement marqué dans le cas de l'apparition de projets spatiaux dans le TTVS, qui permettent de mobiliser des moyens scientifiques et financiers importants propices aux développements.

Exemples : Dans le TTVS, le projet-clé IASI a permis de développer la thématique de l'infrarouge, marquée dans le corpus par l'apparition conjointes de termes tels que *infrarouge*, *infrarouge thermique*, *moyen infrarouge*, *proche infrarouge*, *IR*, *PIR*, *MIR*, *IRT*, etc. Il ne s'agit pas d'une thématique ou de termes/concepts nouveaux, mais le projet IASI permet de réinterroger les possibilités applicables à la détection dans l'infrarouge. Nous avons pu également décrire l'exemple de la thématique des miroirs, rappelée dans le TTVS pour aborder la question des progrès du polissage assisté par ordinateur notamment.

Indices pertinents : Parmi les quatre indices présentés, c'est principalement le regroupement de formes chrono-homogènes (fréquence) qui permet d'accéder à ce type de cas. Les variantes et les dépendances permettent de compléter le repérage.

14. Retrait de thématiques (anciennes mais non obsolètes)

Alors que le démarrage de nouveaux projets entraîne l'apparition de thématiques dans le domaine, l'achèvement de ces projets entraîne quant à lui la disparition de thématiques.

Exemples : C'est le cas notamment de la thématique de l'étalonnage dans le TTVS, qui disparaît du corpus lorsque le projet Spot1 est terminé : *étalonnage en vol, étalonnage au sol, coefficient d'étalonnage, méthodes d'étalonnage, étalonnage radiométrique au sol*

Indices pertinents : Cf. Centralité ponctuelle de thématiques anciennes

15. Modification des connaissances sur un concept donné

Cette catégorie regroupe soit les cas où un concept est modifié, soit les cas où les connaissances des experts sur un concept ont changé (elles ont généralement augmenté), et que nous avons regroupé sous le phénomène de « changement de point de vue ».

Exemples : Les deux exemples décrits dans cette recherche sont ceux d'*opérateur* (DORIS) et *parasites* (TTVS), décrits au Chapitre 7.

Indices pertinents : L'indice qui permet de repérer ces cas est principalement celui des dépendances syntaxiques, lorsque les dépendances se renouvellent et permettent de mettre au jour des changements importants de cooccurrence.

16. Modification de la structure des documents

La dernière évolution pertinente repérée dans cette recherche concerne la documentation technique qui sert de base aux corpus et qui peut également subir des modifications dans le temps. Ainsi, certains termes/concepts apparaissent dans le corpus, non pas parce qu'ils apparaissent dans le domaine, mais parce que la structure ou la forme des documents a pu changer.

Exemples : Dans le projet DORIS, les questions se rapportant au *terminal de transfert*, qui étaient traitées en annexes lors de la première génération, se retrouvent en corps de texte dans la troisième.

Indices pertinents : L'indice de fréquence (termes isolés) semble le seul à permettre

d'identifier ces cas.

17. Mise à jour de la documentation qui constitue le corpus d'étude

Enfin, la dernière catégorie présentée dans cette section regroupe l'ensemble des cas qui ne sont pas pertinents par rapport à la perspective d'évolution des connaissances. Il s'agit de cas où les termes/concepts repérés comme liés à une évolution ne sont en fait liés qu'à la correction ou à la modification des textes les plus récents du corpus. Ceci est particulièrement marqué dans le corpus TTVS, composé de différentes éditions d'un même cours. Ces éditions sont des mises à jour les unes des autres et certains termes/concepts abordent des éléments de connaissances oubliés dans les éditions les plus anciennes, ou encore on corrige la manière d'aborder certaines questions, etc.

Exemples : Thématique des *colles* (TTVS)

Indices pertinents : Essentiellement la fréquence (termes isolés), complétée parfois par les dépendances.

La diversité des évolutions susceptibles d'intervenir en diachronie courte met en jeu différentes facettes de l'évolution, en lien avec les dimensions interne et externe sur lesquelles se base l'analyse. Dans la section suivante, nous soulevons les points principaux liés à ces dimensions dans la mesure où ils permettent de mettre en évidence les questions majeures qui se posent à l'analyste lors du repérage.

8.2 Retour sur les dimensions interne et externe

8.2.1 Dénomination, concept, instance

Nous avons vu dans le Chapitre 2 que l'unité terminologique se définit par le lien qui unit une dénomination et le concept qu'elle dénomme. La dénomination et le concept peuvent toutefois être dissociés, en particulier d'un point de vue diachronique. En effet, comme le remarque Dury (1997), la dénomination et le concept évoluent de manière souvent

indépendante : s'il arrive parfois que dénomination et concept évoluent de concert¹⁰⁹, l'évolution peut intervenir également au niveau de la dénomination seule ou du concept seul. Nous en avons également observé de nombreux exemples dans notre étude. Cet aspect est important dans la mesure où il permet d'éclairer la nature du lien entre dénomination et concept et illustre, à travers un point de vue diachronique, que la dénomination ne peut être considérée comme simple étiquette d'un concept.

L'évolution de l'instance n'est, à notre connaissance, jamais mentionnée en diachronie, mais a été illustrée dans cette étude par les exemples de *Modèle Instrumental* (TTVS) ou *interrupteur secteur* (DORIS). Dans ce cas, nous avons vu que l'entité qui subit le changement n'est ni un concept ni sa dénomination, mais une instance particulière d'un concept dans un contexte précis. Cet aspect éclaire un point important du lien entre dimension interne et dimension externe : bien que le lien entre dénomination et concept soit central (et constitue le point de départ de notre analyse), le parallèle entre langue et connaissances ne doit pas s'y limiter. Ceci confirme la pertinence de considérer le terme comme signe linguistique, de ne pas restreindre la relation interne/externe au couple terme/concept et plutôt de l'élargir à la langue comme système et aux connaissances sous leurs aspects les plus divers.

8.2.2 Différents niveaux d'évolution

Les différentes analyses présentées dans cette recherche montrent effectivement les différents niveaux en jeu dans ce parallèle et la manière dont ils contribuent à marquer une évolution, même sur un intervalle temporel court. Parmi les différents niveaux linguistiques, on peut citer notamment le plan graphique (par ex., on peut noter la suppression du tiret de certains termes complexes (*mini-satellite* → *minisatellite*)), morphologique (par ex. *satellite* → *mini-satellite*), lexical (par ex. *synthèse d'ouverture optique* → *synthèse d'ouverture*), syntaxique (par ex. *communication entre satellites* → *communication intersatellite*), sémantique (par exemple le changement de point de vue sur le concept *parasites*), discursif (par ex. les contextes d'évolution) ou encore lexicométrique (par ex. les empreintes de fréquence).

¹⁰⁹ Dans le cas par exemple de la néologie « complète » (APS, Chapitre 4) ou de certaines variantes qui indiquent la centralité d'un terme/concept dans le domaine (*synthèse d'ouverture optique* qui devient *synthèse d'ouverture*, Chapitre 6). On notera que les cas où dénomination et concept évoluent ensemble ne se résument pas à la néologie seule, mais également à l'implantation ou la centralité.

De la même manière, la variation au niveau de la dimension externe se répartit sur de nombreux aspects. Certaines évolutions concernent bien entendu le niveau conceptuel, tels que les changements de points de vue (Chapitre 7). Au regard de nos observations, et en comparaison avec les travaux existants en diachronie, qui insistent pourtant sur ces phénomènes, on peut dire que l'évolution relève rarement de la néologie (ou de l'obsolescence de concepts) en diachronie courte. Un autre élément externe en jeu dans le cadre de l'évolution des connaissances est celui des instances de concepts. Cet aspect est rarement mentionné mais s'avère remarquable dans nos deux corpus. Parmi les autres phénomènes observables sur la dimension externe, le corpus DORIS a permis de souligner des changements importants dans la structure des documents techniques tels que les rapports de spécification. La structure du cours TTVS a également grandement évolué. On peut faire l'hypothèse que ces changements sont également dus à l'évolution des destinataires pour lesquels ce cours est créé. En effet, le TTVS est un cours conçu à l'attention de jeunes ingénieurs et l'on peut imaginer que les ingénieurs de 1994 sont différents des ingénieurs de 2002 : formation, attentes et connaissances différentes. Ce type d'observation n'est pas trivial pour le linguiste dont l'analyse repose sur le choix de ressources textuelles supposées représentatives pour repérer l'évolution. Ainsi, dans des études telles que celles de Møller (1994) qui compare différentes générations de ressources lexicographiques, on peut penser qu'une proportion – difficilement mesurable cependant – des évolutions repérées est due non pas à l'évolution de la langue, mais à l'évolution de la ressource elle-même (nouvelles perspectives éditoriales, nouvelles contraintes matérielles, etc.). Il ne faut également pas exclure le fait que certains genres textuels sont susceptibles d'évoluer dans le temps (Lozachmeur, 2007 ; Westin & Geisler, 2002). De la même manière, le changement sur la dimension externe peut être marqué par l'influence massive d'un domaine connexe (par ex. l'informatique dans DORIS), la naissance d'un nouveau domaine (par ex. l'optoélectronique, née de l'alliance entre l'optique et l'électronique), l'apparition de nouveaux besoins ou de changements politiques et sociétaux qui viennent modifier les perspectives de recherche (par ex. les accords internationaux pour l'émission des balises DORIS). Au Cnes par exemple, on voit émerger depuis une vingtaine d'années les thématiques du développement durable, de surveillance des changements climatiques, etc.

qui reflètent en grande part des changements politiques et sociaux en France. L'influence de l'Europe est également de plus en plus grande et les perspectives spatiales au Cnes prennent au fil des ans une dimension européenne plus marquée *via* des collaborations plus présentes avec l'Agence Spatiale Européenne (ESA).

Ces différentes évolutions sur les dimensions interne et externe montrent que si le parallèle entre langue et connaissances existe bien, il n'est pas possible de postuler des liens directs et univoques sur la base des indices isolés proposés dans notre recherche. La part de l'interprétation et de l'analyse humaine est centrale, et, comme nous le verrons dans les chapitres suivants, d'autres stratégies de mise en œuvre des indices sont possibles pour renforcer la construction des liens entre évolution dans la langue et évolution des connaissances (cf. chapitres suivants).

Mais avant d'aller plus loin, nous souhaitons revenir sur deux points importants. Le premier concerne la distinction entre technique et science dont la prise en compte permet de mieux saisir le rythme des évolutions. Le second revient sur la nature de l'évolution en tant que phénomène continu, élément qui impose une prise de décision importante de la part de l'analyste.

8.2.3 Évolution des techniques et évolution des sciences

Observer l'évolution des connaissances spécialisées fait nécessairement intervenir la distinction entre technique et science :

« Rappelons ici la distinction traditionnelle entre technique et science – étant entendu qu'il y a toujours à redire sur certaines distinctions conceptuelles trop tranchées et les définitions qui s'ensuivent : la « science » est une connaissance discursive ; elle établit des rapports universels et nécessaires entre les objets d'un langage (logique formelle, mathématiques), entre les phénomènes physiques (sciences physiques, sciences naturelles) – et dans ce cas, elle autorise la prévision – ou entre les faits humains (sciences humaines). La technique (du grec *technikos*, de *techné*, « art ») est un savoir pratique, par

opposition à la science qui est tenue pour un savoir théorique (*épistémé*)¹¹⁰. La technique est « [...] un ensemble d'opérations visant à satisfaire des besoins »¹¹¹. » (Acot, 1999 : 78)

Dans les corpus projet de type DORIS, et de l'avis des experts, c'est la dimension technique qui est la plus présente. Cet aspect s'articule logiquement avec la notion de domaine d'activité telle que définie au Chapitre 3 qui explique la place accordée à la technique dans ce type de corpus. Dans le corpus TTVS, ce type d'observation semble plus délicat à établir, même si encore une fois il semble que plus d'évolutions techniques que scientifiques sont observables.

La tendance marquée de ce type d'observation dans nos corpus peut s'expliquer de deux manières :

1. le domaine spatial est un domaine particulièrement favorable aux évolutions techniques ;
2. parce que, selon Acot (1999 : 78), les techniques évoluent plus rapidement que la science, la diachronie courte permet de capter de manière privilégiée des traces d'évolutions techniques.

D'autres analyses devront venir confirmer ces hypothèses afin de voir quels éléments influent le plus sur ce phénomène pour permettre ainsi de mieux construire l'analyse diachronique en fonction de la distinction science/technique.

Du point de vue conceptuel, la distinction entre science et technique peut être éclairée par celle entre « nominal » et « réel » proposée en philosophie par Ferret (1996 : 74) :

« Nominal et réel doivent se comprendre dans le sens suivant : contrairement aux particuliers naturels dont la science a pour tâche de découvrir *a posteriori* l'essence interne réelle, l'essence des artefacts est nominale dans le sens où c'est nous qui *a priori* décidons entièrement de la fonction attribuée à l'artefact indépendamment de toute nature. »

¹¹⁰ Cf. M.C. Bartholy et P. Acot, *Philosophie, épistémologie, précis de vocabulaire*, Paris, Éditions Magnard, 1976.

¹¹¹ Cf. M.C. Bartholy, J.P. Despin, G. Grandpierre, *La science, épistémologie générale*, Paris, Éditions Magnard, 1978, p71.

Cette distinction renvoie à son tour à la différence entre les notions d'*innovation* et de *découverte*. L'artefact est défini étymologiquement comme « ce qui est réalisé par l'homme », comme un « produit artificiel ». L'homme le crée et décide entièrement de sa nature, sa forme, ses fonctions par rapport à un besoin spécifique. Lorsqu'un nouvel artefact est créé, on parle d'*innovation*. À l'inverse, l'homme n'intervient pas dans la définition des particuliers naturels : ces derniers existent dans le monde, et il doit *découvrir* leur existence, leur fonctionnement, etc. Dans nos corpus, bien que les exemples d'innovations soient plus marqués, ces deux types de phénomènes ont pu être observés. Par exemple, la différence entre le changement de point de vue des concepts *opérateur* et *parasite* illustre cette distinction. Dans le cas d'*opérateur*, il s'agit d'un artefact : l'homme a défini *a priori* l'opérateur et son rôle sur les balises DORIS, « indépendamment de toute nature » et en fonction d'un besoin précis. Dans le cas des *parasites*, il s'agit de particuliers naturels : les scientifiques apprennent à connaître « l'essence interne » de ce concept, qui existe indépendamment de toute décision humaine. Comme décrit dans le Chapitre 7, une des conséquences de cette opposition est que l'on peut analyser le concept d'*opérateur* comme ayant été modifié à travers le temps, alors que dans le cas de *parasites*, c'est la connaissance des experts sur ce concept qui a changé dans le temps et non pas le concept dans son essence.

Enfin, à l'instar de Simondon (2005 : 232), on peut considérer que l'histoire des techniques et celle des sciences entrent dans un cycle potentiellement rapide où l'une et l'autre s'alimentent :

« Les besoins pratiques ont contraint l'homme à améliorer les techniques sans attendre le développement de toutes les sciences ; [...] cette avance est surtout caractérisée par le progrès des adaptations et la relative précarité des auto-corrélations ; c'est en partie ce hiatus au cœur des techniques qui a fait ressentir le besoin de sciences, tandis que les parties déjà constituées des techniques offraient aux sciences naissantes non seulement des problèmes théoriques à résoudre [...], mais encore des moyens d'expérimentation : le besoin d'un complément interne de l'invention déjà adaptée est en fait un appel au savoir scientifique. »

Ces quelques aspects illustrent l'importance, mais aussi la difficulté, de la distinction entre sciences et technique. Il est central néanmoins de la prendre en compte dans l'analyse dans la mesure où elle intervient en diachronie sur plusieurs points : sur la différence entre domaine de connaissance et domaine d'activité, sur la construction de certaines interprétations (cf. les exemples de *parasites* et d'*opérateur* ci-dessus), mais également sur la définition des intervalles et de la périodisation des corpus dans la mesure où les techniques semblent évoluer plus rapidement.

8.2.4 Évolution et continuité

Nous avons souligné à plusieurs reprises dans notre étude (notamment aux chapitres 4, 6 et 7) qu'en dehors des cas d'apparition/de création ou de disparition totale de termes/concepts, l'évolution des connaissances est un phénomène continu. Ce dernier aspect implique que l'analyste doive souvent décider si l'évolution observée est assez marquée dans le domaine ou non pour la prendre en compte. Pour mieux illustrer cet aspect, nous prenons deux exemples de continuité rencontrés dans nos observations, le premier sur la dimension interne, le second sur la dimension externe.

8.2.4.1 Entre discours et langue

D'un point de vue linguistique, un exemple de continuité repose sur la distinction entre langue et discours. Nous l'avons illustré à travers l'observation des variantes et les remarques de certains chercheurs sur le processus de réduction des termes comme phénomène discursif qui, avec le temps, peut donner naissance à de nouvelles unités lexicales (notamment Collet, 1997 ; Jacques, 2003).

Cet aspect est mis en avant par les études sur la néologie, les néologismes étant des phénomènes qui appartiennent à la fois à la langue et au discours (Pruvost & Sablayrolles, 2003 ; Cusin-Berche & Mortureux, 2004) et dont les néologues soulignent le caractère transitoire : un néologisme est une entité transitoire qui vise soit à s'intégrer au système, soit à disparaître (Cusin-Berche & Mortureux, 2004 : 9). C'est ce que notent entre autres Cusin-

Berche et Mortureux (2004 : 9) lorsqu'elles soulignent que l'une des tâches de la linguistique est d'observer « comment le discours fait retour sur la langue, c'est-à-dire sur quelles propriétés linguistiques, et à travers quelles modalités s'opèrent l'accueil des créations discursives et leur intégration au système ». De ce point de vue, on peut dire qu'il existe une continuité entre le discours et la langue : des unités naissent en discours et, avec le temps, s'intègrent dans le système de la langue. Les néologismes peuvent en effet rester au niveau discursif ou finir par s'intégrer au système d'une langue et compléter ainsi son stock lexical. La question cependant reste de savoir définir le moment où un néologisme passe du discours à la langue.

Le même type de question se pose à la terminologie textuelle qui, dans la mesure où elle cherche à extraire les termes et leurs définitions à partir de textes du domaine, questionne justement le passage du niveau discursif à la langue (Condamines, 2003a : 32*sqq.*). La réponse qu'elle apporte repose sur le repérage de régularités linguistiques en discours que l'on peut alors associer à des phénomènes linguistiques dans le système de la langue (*ibidem* ; Rastier, 1991 : 90). En diachronie, l'approche que nous proposons fait appel à une démarche similaire : à partir d'indices linguistiques, on tente de faire émerger des phénomènes réguliers dans les textes dont on fait l'hypothèse qu'ils reflètent des évolutions. L'interprétation et la décision de retenir ces phénomènes reviennent à l'analyste. Par exemple, dans le cas des dépendances syntaxiques, nous avons cherché à observer des changements dans la cooccurrence des termes pour mettre au jour des changements de sens. Or, et nous l'avons vu à travers les exemples de *opérateur* et *parasite*, il peut être difficile de statuer sur le changement de sens ou non (p.243). L'analyste doit savoir juger si les régularités observées sont passagères ou significatives, si elles indiquent bien une évolution stable et avérée, etc. En ce sens, on voit que les régularités observables dans les textes ne sont pas données mais sont le fruit d'une analyse linguistique minutieuse pour proposer une interprétation aussi fiable que possible.

8.2.4.2 Du même à l'autre

Du point de vue externe, la distinction entre continuité et rupture n'est pas triviale et trouve un écho en philosophie, notamment chez Ferret (1996 : 32*sqq.*) qui reprend une dichotomie

des néo-aristotéliens entre deux types de changements, les changements de type I et les changements de type II. Les changements de type I sont ceux qui préservent l'identité d'une entité, tels que le changement de degré, de lieu, de propriété ou de qualité (par exemple le passage du bébé à l'enfant, de l'enfant à l'adulte, de l'adulte au vieillard). Les changements de type II quant à eux entraînent une rupture de l'identité (changement de nature, changement de substance – par exemple le passage du vieillard au cadavre).

Cette distinction¹¹² permet d'aborder un point important quant à l'interprétation de l'évolution : la difficulté de statuer sur le fait qu'un concept (ou une instance) change mais reste le même ou qu'il s'agit d'un nouveau concept, différent. Il peut en effet être difficile de faire le départ entre une évolution qui remplace le concept (type II) et une évolution qui « ne transgresse pas » les limites du concept (type I). C'est ce qu'illustre le paradoxe du Bateau de Thésée :

« Le vaisseau sur lequel Thésée alla et retourna était une galiote à trente rames, que les Athéniens gardèrent jusqu'au temps de Démétrius le Phalérien, en ôtant toujours les vieilles pièces de bois, à mesure qu'elles se pourrissaient, et en y remettant des neuves à leurs places : tellement que depuis, dans les disputes des Philosophes touchant les choses qui s'augmentent, à savoir si elles demeurent une, ou si elles se font autres, cette galiote était toujours alléguée de l'exemple de doute, parce que les uns maintenaient que c'était le même vaisseau, les autres, au contraire, soutenaient que non. » (Plutarque (1951) « Vie de Thésée », *La vie des hommes illustres*, Gallimard, Parus, Bibliothèque de la Pléiade, vol.1, p.21, cité dans (Ferret, 1996 : 17)).

Cette difficulté à décider si le concept ou l'instance sont les mêmes ou sont autres a une répercussion importante sur la question du sentiment d'évolution : la subjectivité du jugement expert sur l'évolution se joue particulièrement sur ces questions où la perception du changement est première. Ceci est difficile à anticiper ou contrôler, bien que, nous allons le voir, certaines possibilités sont envisageables pour guider l'analyste et/ou l'expert dans leurs choix.

¹¹² Bien qu'elle soit appliquée avant tout aux humains et aux particuliers naturels en philosophie, et non pas aux artefacts qui représentent pourtant la majorité des concepts observés dans nos corpus.

8.3 Conséquences méthodologiques

Dans ce chapitre, nous avons dressé un bilan des évolutions interprétables à partir des quatre indices linguistiques que nous proposons, ainsi qu'un bilan des différentes facettes de l'évolution en jeu au cœur du parallèle entre dimensions interne et externe. Ces éléments permettent de mettre au jour des pistes centrales pour établir une méthode de repérage fiable de l'évolution. À la lecture de ces résultats, plusieurs éléments entrent en jeu, éléments qui peuvent influencer les évolutions observées. Parmi ces éléments, nous avons souligné le rôle du type de domaine. En effet, il apparaît qu'un domaine d'activité entraîne majoritairement l'observation d'évolutions techniques et implique d'observer la nouveauté relativement à l'application ou au contexte pour lequel les textes du corpus sont publiés. Cet élément doit être mis en lien avec la question du genre et du degré de spécialisation des textes qui composent le corpus. Comme nous l'avons vu au Chapitre 5, il apparaît que certains indices tels que les contextes riches en connaissances évolutives dépendent du type de corpus en jeu. De la même manière, une piste possible est d'essayer de s'appuyer sur la différence entre évolution des sciences et évolution des techniques pour analyser les données et construire les corpus. Ces éléments guident et influencent les analyses des textes et doivent être pris en compte pour mettre en œuvre de nouvelles études diachroniques. Mais dans la dernière partie de ce travail, nous nous concentrons sur deux autres pistes mises en évidence par nos résultats et qui permettent un retour sur la question de l'interprétation.

Tout d'abord, s'il est clair que les quatre indices permettent bien de repérer de nombreux aspects d'évolution, il reste difficile d'établir une corrélation stable entre un indice et un type d'évolution. En effet, à la lecture des résultats présentés, il est clairement possible d'associer plusieurs types d'interprétations à un seul indice. De la même manière, une même interprétation peut être repérée à l'aide de plusieurs indices. Bien qu'il ne soit pas possible à ce stade de faire émerger des corrélations fiables entre indices et évolution, ceci n'est pas entièrement exclu. Mais dans la mesure où cette piste reste difficilement envisageable pour le moment, une autre solution à explorer est celle de la combinaison d'indices pour repérer l'évolution. En effet, comme le met en évidence le tableau présenté en

Annexe C (p.386), la majorité des interprétations possibles mettent en jeu plusieurs des quatre indices¹¹³. Les combiner laisse envisager de meilleurs résultats pour repérer l'évolution tant du point de vue de la précision du repérage (c'est-à-dire la pertinence des évolutions repérées) que de la qualité des interprétations.

Deuxièmement, le caractère continu de l'évolution nous a permis d'illustrer l'importance du jugement humain pour interpréter l'évolution. L'interprétation humaine est considérée dans notre approche comme partie intégrante du repérage. Cet aspect est cependant loin d'être trivial et l'interprétation doit être guidée au mieux par les corpus et les indices. Nous avons vu jusqu'à maintenant la manière dont les indices permettent de repérer des régularités et des phénomènes pertinents pour l'évolution, mais il est possible d'affiner et de renforcer l'interprétation en combinant les indices, ainsi qu'en prenant en compte dans l'analyse d'autres éléments, tels que l'application pour laquelle on cherche à repérer l'évolution en corpus. Ceci est un point intéressant dans la mesure où habituellement, pour une application donnée, on tente de définir *a priori* des évolutions que l'on souhaite mettre à jour et repérer, sans savoir si ces évolutions sont effectives ou non. Notre démarche est inverse et permet de faire le point sur ce qui est réellement susceptible d'évoluer en diachronie courte et qui est repérable en corpus.

Dans la troisième partie de ce travail, nous développons ces éléments pour construire notre méthode linguistique de repérage de l'évolution en corpus. En effet, maintenant que nous avons montré la pertinence des indices proposés, ainsi que la pertinence d'observer l'évolution en diachronie courte, nous prenons en compte ces nouvelles pistes de réflexion pour développer la démarche d'exploration définie dans notre étude. Pour cela, nous nous concentrons sur la manière de guider l'interprétation pour reconstruire le lien entre indices linguistiques et évolutions extralinguistiques. Les deux derniers chapitres de ce travail développent ces points à travers la description de la possibilité de combiner les indices en faisceaux pour affiner et guider l'interprétation des phénomènes d'évolution, ainsi qu'à

¹¹³ Remarque : dans ce tableau, bien que plusieurs indices puissent être associés à une interprétation, cela ne signifie pas qu'ils sont tous systématiquement « mobilisés » ensemble.

travers une proposition de co-construction de l'interprétation qui intègre l'expert, l'analyste, les corpus et l'application pour parvenir à des analyses fiables et pertinentes.

Troisième partie

Construire l'interprétation : faisceaux d'indices
et principe de co-construction

Chapitre 9 Interprétation de l'évolution *via* des « faisceaux » d'indices

SOMMAIRE DU CHAPITRE

9.1	COMBINER LES INDICES POUR GUIDER L'INTERPRÉTATION.....	275
9.1.1	<i>L'exemple de MOS/CMOS/BiCMOS</i>	276
9.1.2	<i>L'exemple de la thématique du refroidissement</i>	279
9.1.3	<i>L'exemple de l'influence du domaine informatique dans DORIS</i>	281
9.2	ENTRETIENS AUPRÈS D'EXPERTS	285
9.2.1	<i>Entretiens préliminaires</i>	285
9.2.1.1	Tâche proposée	286
9.2.1.2	Résultats.....	288
9.2.2	<i>Seconds entretiens : éléments d'évaluation</i>	291
9.2.2.1	Tâche proposée	291
9.2.2.2	Résultats.....	292
9.2.2.3	Bilan et limites de l'évaluation.....	295
9.2.2.4	Évaluation sur le corpus DORIS.....	297
9.3	BILAN.....	297

Dans la deuxième partie de notre manuscrit, la description des quatre indices montre leur tendance à se combiner pour indiquer un même phénomène d'évolution. Dans ce chapitre, nous revenons sur cette tendance centrale pour l'exploration de l'évolution en corpus et traitons cette question en deux temps. Tout d'abord, à l'aide de trois exemples d'évolution, nous illustrons la manière dont la combinaison d'indices se manifeste en corpus, ainsi que la manière dont elle permet de guider et d'affiner l'interprétation d'un point de vue qualitatif. Dans un second temps, nous rapportons une série d'entretiens auprès d'experts du Cnes rendant compte du fait que la combinaison d'indices est cruciale. Nous montrons premièrement que la combinaison d'indices en faisceaux aide à structurer les entretiens et à guider les experts dans leur interprétation. Deuxièmement, ces entretiens servent de base à l'évaluation des indices et montrent la pertinence de les combiner.

9.1 Combiner les indices pour guider l'interprétation

Dans cette section, nous montrons à travers trois exemples la manière dont la combinaison d'indices en faisceaux permet de guider, compléter et confirmer des interprétations d'évolution à partir d'indices linguistiques en corpus.

9.1.1 L'exemple de MOS/CMOS/BiCMOS

Le terme *MOS* apparaît dans le TTVS et renvoie à deux acceptions (l'une nominale et l'autre adjectivale) :

1. à une technologie de fabrication de composants électroniques
 (lxxx) Une **capacité MOS** est constituée d'un substrat semi-conducteur dopé (pour le silicium type P) recouvert d'un oxyde isolant (silice : SiO₂) sur lequel est disposé un contact métallique servant de grille (TTVS1994)
2. aux transistors fabriqués selon cette technologie
 (lxxxi) Celui-ci est constitué, d'abord par un **MOS**, dont le drain est mis à un potentiel *V_{dr}* (TTVS1998)

Ces deux types d'emplois se rencontrent également pour les termes *CMOS* et *BiCMOS* qui sont aussi deux technologies de fabrication de composants électroniques. Pour ces trois termes, le premier emploi, adjectival, est le plus fréquent et la proportion d'emploi entre ces deux acceptions n'évolue pas au fil du temps : on compte 65% d'emplois adjectivaux pour *MOS* dans l'ensemble des sous-corpus, 90% pour *CMOS* et 75% pour *BiCMOS*. Ces deux acceptions seront donc prises en compte indifféremment dans la suite de l'analyse.

Dans cet exemple, le premier point à souligner est que ces trois termes sont réunis par l'indice de dépendances syntaxiques : le terme *MOS* est productif et « donne naissance » à deux concepts plus spécifiques, *CMOS* et *BiCMOS* (Tableau 9.1).

	TTVS1994	TTVS1998	TTVS2002
<i>MOS (Metal Oxyde Semiconductor)</i>	17	18	20
<i>CMOS (Complementary + MOS)</i>	6	6	7
<i>BiCMOS/bi-CMOS (Bipolar + CMOS)</i>	3	4	4

Tableau 9.1 – Productivité morphologique et lexicale – MOS/CMOS/BiCMOS

Dans la mesure où il est souvent admis que les termes plus spécifiques renvoient à des concepts plus récents dans le domaine, une hypothèse possible est que les *CMOS* sont apparus après les *MOS* et que les *BiCMOS* sont les dernières nées des technologies *MOS*. Les fréquences observables dans le corpus sont très régulières et ne reflètent cependant pas

l'apparition des BiCMOS depuis 1994 (Tableau 9.1) : dès 1994, les trois types de technologies coexistent. Il faut donc nuancer l'hypothèse à la lecture des fréquences et proposer que la productivité permette de repérer un cas de technologies plus récentes (les BiCMOS) mais pas nécessairement nouvelles.

« L'ancienneté » des technologies MOS par rapport aux CMOS et BiCMOS se marque dans nos corpus par deux éléments, signes d'une certaine « maturité » du concept dans le domaine :

- d'abord la capacité du terme MOS à engendrer de nouveaux termes (*BiCMOS*, *CMOS*, *capaMOS*, *photoMOS*, etc.),
- ensuite le fait que les deux acceptions mentionnées plus haut restent assez équilibrées dans les différents sous-corpus (autour de 65% pour l'emploi adjectival et 35% pour la réduction).

D'autres indices peuvent venir renforcer l'hypothèse selon laquelle les BiCMOS sont des technologies récentes entre 1994 et 2002, par exemple la présence de variantes graphiques pour le terme *BiCMOS* (Tableau 9.2). Entre 1994 et 1998, le tiret présent à l'origine disparaît et la graphie de ce terme se stabilise. Comme observé dans le Chapitre 6, l'instabilité graphique peut être un indice de récence d'un terme/concept dans un domaine qui finit son implantation dans le domaine. Et puisque cette stabilisation s'opère entre 1994 et 1998, cela peut renforcer l'hypothèse selon laquelle les BiCMOS sont les technologies MOS les plus récentes, qui ont mûri et s'implantent dans le spatial depuis leur apparition au début des années 90.

	TTVS1994	TTVS1998	TTVS2002
<i>bi-CMOS</i>	3	0	0
<i>BICMOS</i>	0	4	4

Tableau 9.2 – Variantes graphiques bi-CMOS/BiCMOS

Enfin, un dernier indice permet d'appuyer cette interprétation d'évolution, l'existence de contextes riches en connaissances évolutives où ces termes apparaissent :

- (lxxxii) *En ce qui concerne la technologie bi-CMOS, ce n'est qu'en 1993 que nous venons d'avoir accès aux logiciels [...] de conceptions de circuits.* (TTVS1994)
- (lxxxiii) *Aujourd'hui, [...], ce sont les technologies CMOS qui font l'objet d'améliorations incessantes. [...] On constate que les technologies BiCMOS ne sont qu'une amélioration des technologies CMOS existantes.* (TTVS2002)

Ce qui est intéressant dans ce cas est que ces contextes permettent de dater l'émergence des BiCMOS dans le spatial et ainsi de confirmer que ces technologies sont récentes (1993). Ils permettent également de montrer le lien entre les CMOS et les BiCMOS.

Ces contextes montrent enfin qu'il n'y a pas concurrence entre ces technologies : les plus récentes ne viennent pas remplacer les plus anciennes, mais viennent les compléter pour répondre à de nouveaux besoins.

Tout en montrant qu'il est possible de combiner productivité, variantes et contextes, cet exemple illustre l'insuffisance de chacun des indices considérés isolément pour interpréter l'évolution. En effet, si la productivité indique bien une filiation entre les technologies MOS, CMOS et BiCMOS, on ne peut pas dire quand ces dernières sont apparues, ni si elles sont susceptibles de remplacer les plus anciennes ; la stabilisation graphique peut indiquer la récence d'un terme/concept mais pourrait également être due à une correction orthographique lors de la relecture du cours, par rapport à la première édition (puisqu'il ne faut pas oublier que le TTVS1998 est une seconde édition du TTVS et donc une relecture de l'édition précédente) ; enfin, bien que les contextes décrivent explicitement des évolutions, ils doivent être manipulés avec précaution. Dans certains cas en effet, les arguments des rédacteurs du cours ne font pas forcément l'unanimité : si certains présentent une technique particulière comme une clef pour l'avenir du domaine, d'autres peuvent ne pas s'accorder sur ce point. Il faut aussi interpréter avec plus de précaution les marqueurs évoquant une évolution à venir, toujours hypothétique (par exemple *prometteur*).

En d'autres termes, dans ce cas précis, la combinaison d'indices en faisceaux permet à la fois de construire une interprétation fiable (puisque tous ces indices pointent vers la même évolution : l'émergence du concept *BiCMOS* au début des années 1990 dans le spatial et sa filiation aux MOS et CMOS dont il est une amélioration), et d'affiner petit à petit cette

interprétation (la productivité indique la récence et inscrit le terme/concept *BiCMOS* dans sa « famille paradigmatique », les contextes permettent de dater l'émergence de ce concept, d'indiquer qu'ils ne se remplacent pas l'un l'autre, etc.). Le deuxième exemple que nous présentons illustre un cas où la combinaison des indices permet d'affiner l'interprétation à proposer.

9.1.2 L'exemple de la thématique du refroidissement

La thématique du refroidissement émerge dans un premier temps en tant que thématique en évolution dans le TTVS à travers l'indice de fréquence. Dans un premier temps en effet, le regroupement de formes chrono-homogènes met au jour une empreinte de fréquence particulière : la décroissance, à partir de 1998, de termes liés à la thématique de la détection à basse température (§4.3.2.2, p134). 11 termes ont été classés dans cette thématique (présentés par ordre de fréquence) : *froid*, *refroidir*, *luminance*, *bruit*, *signal*, *antenne*, *espace froid*, *dimensionnement radiométrique*, *refroidissement*.

Dans cette liste, seuls sont considérés les termes dont le χ^2 indique une variation de fréquence significative (donc les termes de haute fréquence). Cependant, si l'on observe les termes de basse fréquence, sous réserve de représentativité, on peut ajouter à cette thématique d'autres termes entrant dans la même empreinte et se rapportant à la question du refroidissement pour la détection : *Hélium*, *cryostat*, *bloc froid*, *refroidissement passif*.

Nous l'avons dit, ce type d'empreinte est intéressant dans la mesure où il regroupe près d'une quinzaine de termes thématiquement liés qui présentent une courbe de fréquence atypique : tous diminuent significativement à partir de la deuxième édition du cours.

Cependant, la diminution de fréquence seule ne permet que d'émettre l'hypothèse selon laquelle la question du refroidissement pour la détection à basse température est en décroissance dans le domaine depuis 1998. Mais il reste difficile d'expliquer pourquoi. Il est donc nécessaire d'observer d'autres indices.

Dans cet exemple, les dépendances syntaxiques du verbe *refroidir* (Tableau 9.3) représentent un indice riche d'informations. En effet, elles font apparaître une nouvelle dépendance particulière (en gris dans le tableau) :

	TTVS1994	TTVS1998	TTVS2002
<i>refroidir l'instrument</i>	3	0	0
<i>refroidi par de l'Hélium</i>	2	0	0
<i>refroidi à l'Hélium</i>	2	0	0
<i>non refroidi</i>	0	0	4
<i>refroidir l'optique</i>	0	2	3

Tableau 9.3 – Apparition des nouveaux descendants du terme *refroidir* (TTVS)

Cette nouvelle dépendance (accompagnée par la disparition d'autres qui renvoient au refroidissement) précise la nature d'une nouvelle approche pour le refroidissement : celle du non refroidissement. En contexte, ce « non-refroidissement » s'applique à la fois aux télescopes et aux détecteurs. Le descendant *non refroidi* fait également intervenir dans son champ un terme, jugé néologique par les experts, *microbolomètres* :

- (lxxxiv) Or, récemment, de grands progrès ont été accomplis dans le domaine des mosaïques de détecteurs thermiques **non refroidis** (matrices de **microbolomètres**, par exemple) (TTVS2002)

La néologicit   de ce terme est soulign  e   galement par le type de contextes dans lequel ce terme appara  t : 4 des 5 occurrences de ce terme s'inscrivent dans des contextes d'  volution qui marquent des progr  s et des nouveaut  s tels que :

- (lxxxv) La technologie a justement beaucoup **progress  ** dans ce domaine : les mosa  ques de d  tecteurs    grand nombre de pixels **sont disponibles**, pour les d  tecteurs quantiques (photoconducteurs et photovolta  ques) refroidis, comme pour les d  tecteurs thermiques (microbolom  tres et pyro  lectriques) (TTVS1998 et TTVS2002)
- (lxxxvi) Pour avoir de bonnes performances, ces d  tecteurs doivent en g  n  ral travailler    tr  s basse temp  rature, c'est-  -dire quelques degr  s Kelvin, le plus souvent le dixi  me de degr   Kelvin. Ceci n'est pas vrai pour une **nouvelle** famille de d  tecteurs, les microbolom  tres. (TTVS2002)

Dans cet exemple du refroidissement,    partir de la fr  quence, des d  pendances syntaxiques et des contextes, on peut proposer l'interpr  tation suivante : le refroidissement est une th  matique constante dans l'ensemble du TTVS, mais des progr  s interviennent sur cette

question à partir de 1998. Ces progrès sont liés à de nouvelles techniques et technologies pour ne plus refroidir les détecteurs et les instruments. Les termes/concepts en lien avec la question du refroidissement diminuent donc pour laisser place à des termes/concepts liés au non refroidissement.

À la lecture de contextes et d'occurrences de *non refroidi* tels que :

(lxxxvii) *Par contre, d'autres exemples montrent qu'on peut "vivre" avec un télescope non refroidi (Pronaos embarqué sur une nacelle ballon travaillait aussi dans le domaine infrarouge et submillimétrique). (TTVS2002)*

il s'avère possible de poursuivre encore l'analyse et de lier le développement du non refroidissement à la question de la détection infrarouge. Bien que nous arrêtons là l'analyse de cet exemple, il peut en effet être utile de compléter l'observation par celle d'indices autour de cette thématique de l'infrarouge, question développée en partie dans le Chapitre 6 (§6.3.4, p.214).

Cet exemple illustre donc la manière dont il est possible d'affiner l'interprétation d'une évolution « de fil en aiguille ». En effet, d'une tendance globale dégagée par des regroupements de fréquence, on parvient à tisser – ou plutôt à « détisser » – le fil d'une évolution dans le domaine. La combinaison d'indices en faisceaux est donc un atout méthodologique certain pour repérer et analyser l'évolution des connaissances en diachronie courte.

Cette métaphore de « fil à détisser » est particulièrement présente dans le corpus DORIS, où les évolutions sont peu nombreuses et assez homogènes, comme l'illustre l'exemple suivant.

9.1.3 L'exemple de l'influence du domaine informatique dans DORIS

Le dernier exemple que nous développons concerne la migration de termes/concepts de l'informatique dans DORIS. Le premier indice qui met en évidence la migration de ces termes/concepts est le regroupement de formes chrono-homogènes. Comme illustré au Chapitre 4 (§4.3.2.1, p.132), de nombreux termes/concepts appartenant au domaine informatique sont regroupés comme étant tous en croissance. Plus important encore, la majorité d'entre eux n'apparaît que dans les corpus les plus récents (par ordre de

fréquence) : *bloc, configuration, tch, contexte, télégestion, réception, fichier, recevoir, tlgp, donnée, module de gestion, téléchargement, émission, logiciel, interface.*

Dans un second temps, l'observation peut être recentrée sur chacun de ces termes. Parmi ceux-ci, le terme *interface* révèle des éléments intéressants sur l'influence de l'informatique dans le projet DORIS, en particulier à travers ses dépendances syntaxiques. En effet, ce terme est l'un des seuls de la liste en croissance qui apparaît dans l'ensemble des sous-corpus DORIS, mais dont la distribution change et se complète sensiblement (Tableau 9.4).

	DORISGen1	DORISGen3
<i>interface vers le CDC</i>	2	0
<i>interface CCD</i>	2	0
<i>interface entre capteurs</i>	3	0
<i>spécifications d'interface</i>	4	5
<i>interface télégestion</i>	0	3
<i>interface opérateur</i>	0	12

Tableau 9.4 – Extrait des dépendances syntaxiques d'*interface* (DORIS)

Ces dépendances sont marquées fortement par l'apparition du terme/concept *interface opérateur*, qui compte 12 occurrences dans les textes de troisième génération. L'apparition de ce terme/concept est d'autant plus importante qu'elle est accompagnée de l'apparition simultanée du sigle *IHM* (*interface Homme-Machine*) qui compte 18 occurrences. Ces indices tendent à souligner l'extension de sens que subit le terme *interface* qui, d'élément de jonction entre deux éléments matériels ou logiciels devient élément de jonction entre éléments matériels/logiciels et l'homme.

De ce point de vue, le terme/concept d'*opérateur* apporte lui aussi son lot d'informations : puisque l'informatique permet le développement d'interfaces entre l'homme et la machine, l'opérateur (agent humain chargé d'interagir avec la balise) est susceptible de voir son rôle changer. Et c'est ce que montre le changement de point de vue repérable sur la base des dépendances syntaxiques de ce terme (Tableau 7.15, p.245, repris dans le Tableau 9.5 ci-dessous).

	DORISGen1	DORISGen3
<i>interface opérateur</i>	0	12
<i>opérateur appuyer sur touche</i>	26	0
<i>opérateur appuyer sur val</i>	5	0
<i>opérateur tourner clé</i>	33	0

Tableau 9.5 – Extrait des dépendances syntaxiques d'*opérateur* (DORIS)

Ces dépendances montrent que le rôle de l'opérateur a lui aussi bénéficié de l'influence de l'informatique et que l'ensemble des tâches manuelles qui lui étaient attribuées en tant qu'agent sur la balise sont informatisées, en partie *via* une interface Homme-Machine (appelée *interface opérateur*).

Enfin, il est possible d'extraire des contextes équivalents dans les première et troisième générations de textes qui montrent la modernisation de concepts entraînée par l'informatisation dans DORIS. Ainsi, dans DORISGen1 on peut lire :

(lxxxviii) *La face avant est équipée essentiellement de :*
 - un **afficheur** à cristaux liquide de 80 caractères,
 - un **clavier** 17 touches dont 7 touches de fonction

Et dans DORISGen3 :

(lxxxix) *La balise doit présenter en face avant un **écran de contrôle** et un **clavier**.*

Bien que ces contextes ne constituent pas des contextes riches en connaissances évolutives tels que décrits au Chapitre 5, ils illustrent cependant le fait que le concept d'*afficheur* semble devenu obsolète au profit du concept *écran*, sur la base duquel se fait le lien entre l'opérateur et les données qu'il doit traiter. De la même manière, si en 1990 on prenait soin de décrire la composition du clavier, ceci n'est plus utile en 2000 où le concept *clavier* est bien harmonisé et familier de tout un chacun.

Dans cet exemple, l'ensemble des indices décrits montre la manière dont la migration de connaissances informatiques vient bouleverser les connaissances d'un projet. La combinaison d'indices dans ce cas sert à repérer ces différentes conséquences. On voit à cette occasion que ce que l'on peut considérer comme une évolution (la migration de

connaissances informatiques) peut se révéler également vectrice, ou cause, d'autres évolutions que seule la combinaison d'indices permet d'identifier.

Ces trois exemples, parmi l'ensemble des analyses possibles dans nos corpus¹¹⁴, permettent d'illustrer le fait que la combinaison d'indices en faisceaux guide l'interprétation (pour ne pas dire l'enquête) de l'analyste et ce, sur au moins trois aspects :

- la combinaison permet d'assurer une interprétation qu'un indice isolé rendrait trop faible ou discutable (*MOS, CMOS, BiCMOS*),
- elle permet d'affiner l'interprétation pour circonscrire avec précision le phénomène en jeu (thématique du refroidissement),
- elle permet de mettre au jour un ensemble de changements liés à une même évolution dont ils sont des conséquences (migration de connaissances informatiques et ses répercussions dans le projet DORIS).

Dans la mesure du possible, il est donc important de chercher à privilégier des « faisceaux d'indices » pour postuler une évolution potentielle dans le domaine. Toutefois, la validation finale de l'évolution doit prendre en compte l'avis d'un expert.

La seconde section de ce chapitre rapporte une série d'entretiens menés auprès d'experts du Cnes. Plus précisément, ces entretiens sont rapportés ici pour aborder la question de la combinaison d'indices en faisceaux sous deux angles :

1. comme une étape méthodologique nécessaire pour construire des tâches pertinentes à proposer aux experts ;
2. comme moyen d'évaluer la pertinence des indices et de leur combinaison.

¹¹⁴ Notamment concernant les thématiques des lasers, des miroirs et de l'infrarouge, mentionnées à plusieurs reprises dans les descriptions rapportées dans notre manuscrit.

9.2 Entretiens auprès d'experts¹¹⁵

9.2.1 Entretiens préliminaires

Avant de décrire les entretiens mis en place, il convient de présenter plusieurs remarques relatives à l'importance et à la difficulté d'une telle tâche. La première difficulté est liée au fait que, bien que le recours aux experts soit très fréquent en terminologie, très peu de travaux détaillent ces entretiens ou proposent des retours d'expérience qui permettraient de définir des tâches de manière éclairée et contrôlée. Et, la perspective diachronique étant le parent pauvre des analyses terminologiques, le manque de retours d'expérience y est d'autant plus sensible. Les seuls éléments dont on dispose à ce jour sont les travaux de néologues qui se sont penchés sur la question du sentiment néologique des locuteurs (en particulier Ben Hariz Ouenniche, à paraître ; Gardin, *et al.*, 1974 ; Sablayrolles, 2003a). Dans ces travaux, les auteurs interrogent des locuteurs sur leur sentiment face à la néologité de mots et de ce fait amorcent une réflexion sur le recours à des informateurs extérieurs pour travailler sur l'évolution.

Ces travaux sont cependant importants dans la mesure où ils permettent de souligner l'instabilité et la subjectivité du sentiment d'évolution des locuteurs ainsi que la difficulté pour eux d'établir un « rétro-diagnostic » (Dury, 2008a). Nous y reviendrons en détail dans le chapitre suivant (§10.2.2.3, p.307), mais il est clair en effet qu'interroger un individu sur son sentiment face à l'évolution (que ce soit de la langue ou des connaissances) est un acte délicat qu'il est nécessaire de guider au mieux. À ceci s'ajoute le fait que la mise en place d'entretiens entre analyste et expert repose sur la relation entre ces deux partis. Cette question n'est quant à elle jamais abordée dans la littérature mais est centrale, en particulier en diachronie, dans la mesure où interroger un expert sur l'évolution d'un domaine revient à l'interroger sur ce qui varie et qui peut échapper à son « contrôle ». De fait, interroger l'expert sur ce qui évolue peut vite avoir pour conséquence de remettre en question la représentation que ce dernier a de son domaine. Ce sentiment est d'autant plus probable que l'entretien est mené par un linguiste ou un terminologue qui, qu'il le veuille ou non, représente souvent le juge de la compétence linguistique de l'expert, compétence qui

¹¹⁵ Les résultats présentés dans cette section ont été publiés dans (Picton, 2008).

englobe sa grammaire, son vocabulaire et, par extension, tout type de production textuelle ou orale dont il est l'auteur, dont les corpus utilisés pour l'analyse. Dans la perspective diachronique, ceci se traduit en pratique par des résultats d'entretiens partiels ou caducs où l'expert se refuse à admettre certaines évolutions. Il est donc souvent nécessaire lors de l'entretien de « désamorcer » ce sentiment, en dialoguant avec l'expert et en choisissant une consigne qui n'accentue pas cette variation. Il est important pour l'analyste d'apprendre à ajuster la formulation de la consigne entre des tâches trop génériques qui risquent de perdre l'expert et de l'éloigner des objectifs de l'étude (mais qui offrent aussi l'avantage de lui laisser plus de liberté sur les informations qu'il peut dévoiler) et des consignes trop spécifiques qui « bloquent » son jugement (mais qui du point de vue de l'analyste lui permettent de cibler l'analyse d'un élément spécifique). Dans la mesure du possible, il est donc important de rencontrer au préalable le ou les expert(s) afin d'établir le dialogue et de mieux comprendre à la fois les attentes de l'expert et de l'analyste face à la tâche envisagée.

De ce point de vue, la formulation de la consigne devient essentielle pour obtenir des résultats fiables (Devereaux Ferguson, 2000 : 147). Celle-ci doit permettre à la fois de répondre aux attentes de l'analyste et d'être immédiatement compréhensible par l'expert. Comme le souligne Josselin-Leray (2005 : 280) qui reprend Jacquart (1988 : 75-76) à propos de la mise en place de sondages, il est en particulier nécessaire de « se mettre à la portée de la personne interrogée, en utilisant des termes simples ». Et lorsqu'il s'agit de présenter à l'expert une liste de termes, la combinaison des indices permet justement d'ajuster la tâche. C'est ce que nous illustrons dans le paragraphe suivant.

9.2.1.1 Tâche proposée

Dans notre recherche, plusieurs rencontres ont pu être mises en place, dont les premières ont pu jouer un rôle de « prise de contact » avec deux de nos 4 experts d'optique spatiale. Ces rencontres ont effectivement révélé le décalage entre nos attentes et celles des experts. Notamment, elles ont confirmé l'instabilité du sentiment d'évolution des experts, non seulement entre experts, mais également chez un même expert¹¹⁶. Pour cette raison, les

¹¹⁶ Nous revenons sur ce point dans le chapitre suivant.

consignes ont souvent été contraintes à un seul aspect d'évolution à la fois et sont établies pour laisser le champ libre à la discussion avec l'expert. Ces discussions lui permettent d'affiner et de nuancer son sentiment face à l'évolution des termes/concepts présentés. Nous espérons ainsi créer au mieux l'équilibre entre contrainte et ciblage mentionné *supra*.

Dans les entretiens rapportés dans ce chapitre, la consigne établie est la suivante :

Observez cette liste de termes. Ceux-ci n'apparaissent pas dans la première édition du cours (TTVS1994) mais apparaissent dans la dernière (TTVS 2002). À votre avis, peut-on dire qu'ils renvoient à des nouveautés en optique spatiale ? Ou peuvent-ils être mis en lien avec des progrès/évolutions dans ce domaine ? Dans le cas contraire, précisez.

Précisons simplement que, par rapport à ce qui a été décrit dans le Chapitre 4, les termes retenus dans cette liste ne se limitent pas aux seuls termes dont le χ^2 est significatif. Elle contient donc dans un premier temps les 278 termes apparus dans le TTVS (§4.3.1.1, p.115). Ceci s'explique pour deux raisons. La première est que la mise en place de plusieurs entretiens avec plusieurs experts est une entreprise de longue haleine qui s'est étalée pour cette recherche entre juin 2006 et décembre 2008 ; de fait une partie des entretiens a été mise en place avant le choix définitif du filtre de χ^2 pour sélectionner les empreintes de fréquence. La seconde raison nous permettant de maintenir ce choix est que le but de ces entretiens préliminaires est d'ajuster la tâche. Il est nécessaire alors de n'avoir aucun parti pris sur la pertinence des termes en fonction de leur fréquence.

La liste présentée aux experts pour cette tâche contient des termes de fréquence nulle dans le TTVS1994 et une fréquence non nulle dans le TTVS2002. Nous l'avons vu à plusieurs reprises dans notre travail, l'indice de fréquence est un indice intéressant pour commencer à explorer l'évolution dans la mesure où il permet d'associer des hypothèses d'évolution assez claires et systématiques. C'est le cas notamment de la nouveauté, sur laquelle notre consigne est basée.

Cette liste de 278 termes est présentée par ordre de fréquence (Figure 9.1), du terme le plus fréquent au terme le moins fréquent, mais les experts ont le choix de reclasser différemment

cette liste, notamment par ordre alphabétique, choix que la majorité des experts impliqués dans l'ensemble des tâches menées ont fait.

bilan radiatif
dispositif d'étalonnage
filtre passe-bande
flux lumineux parasite
imagerie thermique
laser à cavité étendue

Figure 9.1 – Extrait de la liste présentée (premiers entretiens)

9.2.1.2 Résultats

Pour mémoire, la tâche proposée repose sur deux questions :

1. la nouveauté des termes/concepts présentés,
2. sinon, leur lien éventuel avec des évolutions dans le domaine.

Sur ces 278 termes, seuls 46 ont pu être traités par les experts. La raison principale de ce faible nombre est que la tâche de validation de l'évolution de termes est une tâche qui est longue et fastidieuse. Le jugement est difficile à porter, mérite souvent réflexion et, parfois, implique la consultation d'autres informations telles que des contextes d'apparition des termes, ce qui demande du temps. La validation passe donc souvent par de longues périodes de discussion et d'échange entre l'analyste et l'expert. Comme nous l'avons dit, ces périodes, bien qu'elles puissent ralentir la validation, se révèlent d'un apport précieux pour l'analyste puisqu'il s'agit d'une des situations où les experts se livrent le plus et peuvent donner des informations très riches quant à l'évolution de leur domaine.

Parmi les 46 termes/concepts traités, les experts ont validé et jugé nouveaux seulement 5 termes (*APS*, *minisatellite*, *laser à cavité étendue*, *microbolomètre* et *amplificateur à fibre dopée*), soit 10,9% des termes/concepts. Cette faible proportion de termes validés en tant que nouveaux a été surprenante au début de notre recherche mais montre clairement les limites de l'indice de fréquence utilisé seul dans un corpus tel que le TTVS. Ce constat fait, deux autres interprétations principales ont été cependant proposées par les experts pour expliquer la présence des 41 autres termes/concepts dans la liste :

1. **Les termes/concepts sont liés à un progrès, une évolution** : cette catégorie est large et recouvre l'ensemble des termes/concepts que les experts lient à une évolution sans qu'il s'agisse de nouveauté. Elle correspond à la deuxième question formulée dans la consigne. À ce stade, il est difficile pour les experts d'affiner davantage leur jugement. Cette catégorie est donc conservée telle quelle pour les autres tâches et les termes qu'elle contient sont potentiellement pertinents.
2. **Les termes sont liés à un enrichissement du cours et non pas à une évolution du domaine** : cette catégorie regroupe les termes non pertinents repérés par la fréquence. Il s'agit de termes dont l'apparition n'est due qu'à un enrichissement du cours qui sert de base au corpus. Comme nous l'avons dit, ce bruit est lié en grande part à la nature du corpus TTVS puisqu'il s'agit d'un cours composé de trois éditions successives.

Dans la mesure où l'objectif ici est de repérer la nouveauté et l'évolution, il est nécessaire de parvenir à distinguer ce qui relève réellement de l'évolution (catégories 1 et 2) et ce qui relève de l'enrichissement du corpus (catégorie 3). De plus, le fait que cette première liste contienne une si grande proportion de termes non pertinents souligne les limites de l'indice de fréquence considéré seul, puisque moins de 11 % de termes/concepts pertinents sont extraits grâce à cet indice. Il faut donc envisager d'autres stratégies pour améliorer l'extraction des connaissances évolutives.

De plus, il est nécessaire de réviser cette liste pour poursuivre l'évaluation auprès d'experts. En effet, le fait de présenter une liste trop « bruitée » peut entraîner un sentiment de doute, voire de découragement, de la part de l'analyste et de l'expert. En effet, la mauvaise qualité de la liste présentée au regard de la tâche demandée peut mener l'expert à s'interroger sur la pertinence du travail du linguiste, ainsi que sur la pertinence de son propre investissement. L'analyste quant à lui peut également douter de sa démarche et le dialogue avec l'expert risque de tourner court.

Présenter une liste de mauvaise qualité à l'expert entraîne également une autre conséquence importante : comme nous l'avons souligné, la difficulté lorsque l'on interroge un tiers sur sa compétence dans le temps est de devoir l'interroger sur ce qui varie. De fait, pour éviter

l'instabilité que peut provoquer ce genre de tâche, la solution de facilité pour l'expert est d'identifier d'abord ce qu'il juge ne pas avoir évolué. Dans une liste trop bruitée, l'expert est mené à lister tellement de termes/concepts non évolutifs qu'il risque de classer comme non pertinents des termes/concepts qui ont évolué mais sur lesquels son sentiment d'évolution est biaisé par une liste de trop mauvaise qualité.

Pour toutes ces raisons, il est donc nécessaire de filtrer de nouveau ce type de liste. L'hypothèse ici est que les résultats seront meilleurs si l'on combine deux indices pour extraire des termes/concepts qui ont évolué. Pour cela, la liste a été filtrée à l'aide d'un deuxième indice, les dépendances syntaxiques, afin de ne garder que les termes entretenant une relation de dépendance entre eux. Notre choix s'est porté sur cet indice pour une raison principale : lors de ces premiers entretiens, la liste était présentée par défaut aux experts par ordre de fréquence. Ces derniers avaient néanmoins la possibilité de changer cet ordre et d'adopter une classification des termes par ordre alphabétique, ce que les deux experts interrogés ont choisi de faire (cf. *supra*). Nous avons ainsi pu remarquer que travailler avec des termes liés graphiquement (et pour certains morphologiquement et syntaxiquement) rendait la tâche plus facile aux experts. Le fait que certains termes soient regroupés « par famille » semble en effet leur permettre de poser un jugement non pas sur des termes isolés, mais sur des ensembles de termes liés. On peut penser que l'expert adopte alors une réflexion plus générale sur l'évolution d'un concept du domaine, ce qui l'aide à porter plus facilement un jugement sur les concepts appartenant à cet ensemble.

Sur cette base, la seconde série d'entretiens rapportés peut être envisagée comme une étape d'évaluation de la pertinence des indices et de leurs combinaisons en faisceaux pour améliorer le repérage de l'évolution.

9.2.2 Seconds entretiens : éléments d'évaluation

9.2.2.1 Tâche proposée

Pour ces entretiens, la consigne donnée aux experts est la même que précédemment. Pour cette tâche trois experts ont été interrogés, dont un déjà lors des entretiens précédents¹¹⁷.

La liste proposée contient 75 termes (Figure 9.2) :

- dont la fréquence est égale à 0 dans le premier sous-corpus et est non nulle dans le plus récent ;
- et dont la tête est présente dans les trois sous-corpus (dépendances syntaxiques).

Ainsi, comme illustré à la Figure 9.2, les termes *fibre dopée*, *fibre monomode*, *fibre multimode* sont présentés dans la liste dans la mesure où ils sont absents du TTVS1994 mais présents dans le TTVS2002. Leur tête, *fibre*, est quant à elle présente dans les trois sous-corpus. Rappelons qu'elle n'est pas présentée dans cette liste.

(fibre)	<i>fibre dopée</i>
	<i>fibre monomode</i>
	<i>fibre multimode</i>
(laser)	<i>laser à cavité étendue</i>
	<i>laser à fibre dopée</i>
	<i>laser semi-conducteur</i>
(miroir)	<i>miroir plan</i>
	<i>miroir sélectif</i>
	<i>miroir sphérique</i>

Figure 9.2 – Extrait de la liste présentée (seconds entretiens)

Dans cette liste de 75 termes, les experts en ont éliminé 5 à l'unanimité : *asic ciselé*, *dispersion des caractéristiques électriques*, *électronique associée*, *bloc de base* et *surface optique*. Ces termes étaient soit :

- inconnus des experts (*asic ciselé*),
- jugés trop génériques (*bloc de base*, *surface optique*),
- considérés comme n'étant pas des termes de l'optique (*dispersion des caractéristiques électriques*, *électronique associée*).

¹¹⁷ Comme nous le verrons dans le chapitre suivant, le fait de pouvoir interroger plusieurs experts offre de plus grandes chances de parvenir à un consensus sur les évolutions observées.

3 autres termes ont également été exclus de l'analyse : *bandes étroites*, *spectromètre à prisme* et *télescope de Korch*. Les experts ont été en effet en désaccord quant à leur évolution. La liste finale prise en compte dans l'analyse qui suit contient donc 67 termes.

Notons que ces premiers éléments montrent d'emblée que cette deuxième liste, filtrée, est plus facile à analyser pour les experts qui peuvent non seulement traiter tous les termes, mais osent également proposer des jugements sur leur appartenance au domaine ou non.

9.2.2.2 Résultats

Le Tableau 9.6 présente la répartition des 67 termes entre les trois types d'interprétations proposés par les experts lors des premiers entretiens. La deuxième colonne indique le nombre de termes concernés par chaque interprétation.

Jugement	Nombre de termes/concepts
Nouveauté	11
Évolution (progrès, amélioration, projet, etc.)	22
Enrichissement du cours	34
Total	67

Tableau 9.6 – Jugement des experts sur la seconde liste

Comme souligné *supra*, les aspects qui nous intéressent concernent aussi bien la nouveauté que les progrès, améliorations, etc. Les termes pertinents sont donc les 33 termes répartis dans les lignes 2 et 3 en gris.

D'emblée, ces résultats montrent clairement que filtrer les résultats avec les indices de fréquence et de dépendances associés entraîne une meilleure sélection des résultats : 33 termes sur 67 sont jugés pertinents, soit un peu plus de 49%. Rappelons que lors des premiers entretiens, l'indice de fréquence seul avait permis d'obtenir 10% de candidats pertinents.

Pour compléter l'évaluation, il est également intéressant d'observer les autres termes de cette liste concernés par les deux autres indices d'évolution décrits dans cette étude, à savoir les variantes et les contextes.

Parmi les 67 termes de la liste finale, 20 ont des variantes et 24 apparaissent dans des contextes riches en connaissances évolutives. À l'aide de ces éléments, il devient possible de répondre à au moins deux questions :

1. Quelle est la proportion de termes (évolutifs) concernés par les indices linguistiques identifiés ?
2. Dans quelle mesure la combinaison de plusieurs de ces indices permet-elle d'améliorer le repérage de l'évolution ?

Le Tableau 9.7 donne quelques éléments pour répondre à la première question.

Nombre d'indices considérés	Nombre de termes extraits	Proportion de termes extraits
<i>Au moins 2 indices</i>	67	100 %
<i>Au moins 3 indices</i>	32	47,76 %
<i>4 indices</i>	12	17,91 %

Tableau 9.7 – Nombre et proportion de termes concernés par les indices

Dans ce tableau, il apparaît clairement que plus on combine d'indices, moins il y a de termes/concepts concernés : près de 50 % des termes/concepts sont concernés par trois indices et une minorité (moins de 20 %) par les quatre indices¹¹⁸.

Au niveau de l'apport des combinaisons d'indices pour extraire des termes/concepts évolutifs pertinents, la tendance inverse est observée : plus on prend d'indices en compte, plus l'extraction est précise et performante (Tableau 9.8).

¹¹⁸ Rappelons que les termes de la liste sont tous concernés par au moins deux indices puisque la liste est filtrée à l'aide de la fréquence et des dépendances syntaxiques.

Nombre d'indices considérés	Nombre de termes <u>pertinents</u> extraits	Proportion de termes <u>pertinents</u> extraits
<i>Au moins 2 indices</i>	33	49,25 %
<i>Au moins 3 indices</i>	22	68,75 %
<i>4 indices</i>	10	83,33 %

Tableau 9.8 – Nombre et proportion de termes pertinents concernés par les indices

Ainsi, si l'on extrait un peu moins de 50 % de termes/concepts pertinents en combinant deux indices, on en extrait près de 70 % en combinant trois indices et plus de 80 % en combinant les quatre indices. Ces résultats montrent que combiner les indices est une piste intéressante pour extraire les termes évolutifs pertinents de manière satisfaisante. Néanmoins, comme peu de termes sont concernés par les quatre indices simultanément, bien que l'extraction soit précise, beaucoup de termes/concepts évolutifs peuvent être manqués. À ce stade, il faut donc faire un choix entre la quantité et la qualité des informations extraites. Ce choix peut être guidé par le contexte et l'application dans lesquels prend place l'étude (Chapitre 10).

D'un point de vue qualitatif, on peut observer certains liens entre les indices et des aspects spécifiques d'évolution. Comme l'illustre le Tableau 9.6, les experts ont jugé que 11 candidats sont liés à de la nouveauté et 22 à une évolution telle que des progrès, des améliorations, etc. Si l'on observe les 11 termes nouveaux, on remarque que 10 d'entre eux peuvent être repérés par la combinaison des quatre indices. Bien qu'il faille confirmer cet aspect par d'autres expérimentations à plus grande échelle, cela constitue d'emblée une piste intéressante quant au traitement de la nouveauté (et donc de la néologie), qui semble l'aspect le plus marqué dans les textes.

Enfin, nous terminons sur des remarques concernant les cas de combinaison de 3 indices. Dans cette configuration, et à partir de notre liste, deux combinaisons d'indices sont possibles :

1. la fréquence, les dépendances et les variantes,
2. la fréquence, les dépendances et les contextes.

Les Tableau 9.9 et Tableau 9.10 présentent une évaluation comparée de ces deux combinaisons :

Nombre d'indices considérés	Nombre de termes extraits	Proportion de termes extraits
Variantes	8	11,94 %
Contextes	12	17,91 %

Tableau 9.9 – Nombre et proportion de termes concernés par les variantes vs. les contextes

Nombre d'indices considérés	Nombre de termes <u>pertinents</u> extraits	Proportion de termes <u>pertinents</u> extraits
Variantes	5	62,50 %
Contextes	7	58,33 %

Tableau 9.10 – Nombre et proportion de termes pertinents concernés par les variantes vs. les contextes

Ces résultats montrent que les contextes sont un peu plus fréquents que les variantes, mais que leur précision est comparable (autour de 60 %). D'un point de vue qualitatif, on observe cependant une différence importante entre ces deux indices : les 5 termes pertinents concernés par les variantes sont tous classés dans la catégorie des évolutions du type progrès, améliorations, etc. alors que les 7 termes pertinents concernés par les marqueurs et contextes sont liés à de la nouveauté. Encore une fois, ce phénomène mériterait d'être vérifié par d'autres expérimentations, mais il semble que dans cette liste, les contextes sont un indice plus adapté pour repérer la nouveauté¹¹⁹.

9.2.2.3 Bilan et limites de l'évaluation

Ces éléments d'évaluation sont très riches et offrent des pistes précieuses pour évaluer l'intérêt des indices et de leurs combinaisons. Néanmoins, ces seuls entretiens restent insuffisants. En effet, pour que l'évaluation soit satisfaisante, il faudrait pouvoir répondre à au moins deux questions supplémentaires :

1. Certains indices sont-ils plus efficaces ?
2. Quelles sont les autres combinaisons possibles et leurs performances respectives ?

¹¹⁹ Sous réserve que cela ne soit pas lié uniquement à la liste de marqueurs que nous avons définie.

Malheureusement, répondre à ces questions est une tâche difficile à laquelle notre contexte d'évaluation auprès d'experts n'est pas parfaitement adapté.

Le recours aux experts comme mode d'évaluation est un choix coûteux en temps. Si cette démarche garantit une évaluation riche d'informations, un inconvénient majeur est que toutes les combinaisons d'indices n'ont pu être testées faute de temps. En effet, avec quatre indices 16 combinaisons comportant de un à quatre indices sont envisageables. Dans ce chapitre nous n'en testons que 5 :

- 1 indice seul : la fréquence,
- 1 combinaison de 2 indices : la fréquence et les dépendances
- 2 combinaisons de 3 indices :
 - a. la fréquence, les dépendances et les variantes
 - b. la fréquence, les dépendances et les contextes
- la combinaison des 4 indices : la fréquence, les dépendances, les variantes, les contextes.

De plus, une limite importante est qu'il existe entre les indices certains liens de dépendance modifiant les résultats. En particulier, la fréquence est un indice qui vient modifier sensiblement la lecture que l'on peut faire des trois autres. En effet, il n'est par exemple possible de prêter une interprétation diachronique aux variantes et à la productivité que si l'indice de fréquence est vérifié (cf. Chapitres 6 et 7). Il sera donc important d'essayer d'évaluer la précision des différentes combinaisons et des indices isolés pour mesurer l'influence des indices entre eux.

Une autre limite à cette évaluation est qu'il est difficile de la reproduire à l'identique, dans la mesure où elle repose sur une collaboration étroite avec les experts (ce qui implique une part de subjectivité), et dans la mesure où elle peut dépendre du type de corpus. De fait, les résultats présentés sont un peu différents pour le corpus DORIS, pour lequel nous n'avons collaboré qu'avec un expert.

9.2.2.4 Évaluation sur le corpus DORIS

Le même type d'expérience et d'entretien a pu être mené pour le corpus DORIS, mais diffère de celui mis en place pour le TTVS sur deux plans principaux. Le premier, comme nous l'avons dit, est que pour DORIS nous collaborons avec un seul expert. Le second est que les indices d'évolution dépendent étroitement du genre et du degré de spécialisation du corpus. Ainsi, comme décrit au Chapitre 5, les contextes riches en connaissances évolutives sont quasiment absents du corpus DORIS et n'ont pu être pris en compte.

Néanmoins, nous avons mené des évaluations préliminaires sur le même type de tâche avec une liste de 60 candidats extraits à l'aide de l'indice de fréquence seul. Dans cette première liste, notre expert a jugé qu'un peu plus de 36% des termes extraits étaient pertinents (soit 22 termes/concepts).

Sur une deuxième liste de 60 candidats, extraits cette fois-ci à l'aide de l'indice de fréquence et de la productivité, il a jugé que 64% des termes extraits étaient pertinents (soit 38 termes/concepts). Bien qu'incomplète, cette évaluation est intéressante dans la mesure où elle corrobore les résultats obtenus dans le corpus TTVS : plus on combine d'indices et plus l'extraction est précise.

Une autre piste intéressante concerne le lien entre type d'indice et type de corpus. En effet, bien que la combinaison d'indices présente la même tendance entre les deux corpus, on s'aperçoit que la fréquence, considérée seule, entraîne de meilleurs résultats d'extraction dans DORIS. Cette remarque fait écho à nos observations dans la deuxième partie de notre travail, où nous avons souligné l'homogénéité des évolutions dans DORIS par rapport au TTVS.

9.3 Bilan

Les deux points de vue présentés ici sur la combinaison d'indices en faisceaux montrent l'intérêt méthodologique de ce type de démarche sur plusieurs points.

Tout d'abord, combiner les indices permet d'améliorer la pertinence des termes/concepts extraits. Bien que cela implique de ne pas repérer tous les termes/concepts potentiellement évolutifs, ceci laisse le choix à l'analyste quant aux listes qu'il souhaite traiter.

Dans le cas où il collabore avec des experts, et c'est le deuxième point, la combinaison permet à l'analyste de guider l'expert dans ses observations et de proposer une liste et une consigne qui favorisent le dialogue entre eux.

Néanmoins, à l'issue d'entretiens tels que ceux rapportés dans la section 9.2, des évolutions pertinentes sont mises au jour mais ne restent qu'assez grossièrement caractérisées. Afin d'affiner l'interprétation des évolutions en jeu, la combinaison d'indices est une nouvelle fois une piste importante. Comme l'illustrent les trois exemples proposés à la section 9.1, la combinaison d'indices s'avère riche d'informations pour affiner, confirmer et compléter les interprétations sur l'évolution des connaissances. Ceci est d'autant plus important que, comme nous l'avons vu au Chapitre 8, les évolutions repérables en diachronie courte sont très hétérogènes, mais également très fines. En effet, les phénomènes d'évolution qui interviennent sur de courtes périodes se manifestent souvent à un grain très fin, souvent plus fin qu'en diachronie longue.

Ce travail d'enquête n'est cependant possible que dans la mesure où l'on envisage notre méthodologie comme la construction d'une interprétation, construction menée par de multiples acteurs. Nous postulons en effet que, pour parvenir à des interprétations des phénomènes d'évolution, le travail en corpus s'inscrit dans une co-construction de l'interprétation des données, co-construction qui fait intervenir l'analyste, l'expert, les indices et les corpus, mais également l'application pour laquelle le repérage est mis en place.

Chapitre 10 Pour une co-construction de l'interprétation

SOMMAIRE DU CHAPITRE

10.1	PRINCIPE DE CO-CONSTRUCTION	300
10.1.1	<i>Construire l'interprétation</i>	300
10.1.2	<i>Consensus interprétatif</i>	301
10.2	ACTEURS DE LA CO-CONSTRUCTION.....	303
10.2.1	<i>Analyste</i>	303
10.2.1.1	Définition de l'analyste.....	303
10.2.1.2	Atouts : l'analyse de corpus de textes.....	303
10.2.1.3	Limites de l'analyste.....	304
10.2.2	<i>Expert du domaine</i>	305
10.2.2.1	Rôle de l'expert.....	305
10.2.2.2	Qualités et atouts d'un expert pour la diachronie.....	306
10.2.2.3	Difficultés du « rétro-diagnostic »	307
10.2.3	<i>Nature des échanges</i>	313
10.3	PRISE EN COMPTE DE L'APPLICATION	314
10.3.1	<i>Application comme guide de l'analyse</i>	314
10.3.2	<i>L'exemple de la mise à jour de ressources termino-ontologiques</i>	316

La méthodologie proposée dans cette étude repose sur une analyse linguistique en corpus. Les résultats présentés dans ce travail nous poussent à proposer des arguments en faveur d'une co-construction de l'interprétation des phénomènes de l'évolution en corpus. Cette proposition s'appuie sur le fait que les phénomènes observables ne sont pas donnés par les corpus, mais doivent faire l'objet d'une interprétation (Bachelard, 1938 ; Rastier, 2001). Plus précisément, en diachronie l'interprétation est le fruit d'une construction *conjointe* de la part d'un analyste et d'experts du domaine qui s'appuient sur des textes, des indices linguistiques et une application.

Dans ce chapitre, la première section développe le principe de co-construction proposé pour notre démarche. La deuxième section présente les deux acteurs de la co-construction et permet entre autres d'aborder la question du rôle et de la place des experts à la lueur de la perspective diachronique. La troisième section montre l'importance de prendre en compte l'application pour mener l'analyse et la manière dont celle-ci peut guider les choix méthodologiques que nous avons posés dans notre recherche.

10.1 Principe de co-construction

10.1.1 Construire l'interprétation

Le recours aux corpus de textes est une approche souvent incontournable en diachronie (Chapitre 3). Cette perspective est très riche et prometteuse d'un point de vue méthodologique, et nous avons vu dans l'ensemble de cette étude que le lien entre dimension interne et dimension externe existe bel et bien en diachronie courte. Celui-ci s'est révélé cependant très hétérogène dans la mesure où à un phénomène (indice) sur la dimension interne peuvent correspondre de nombreux phénomènes sur la dimension externe (Chapitre 8). Dans notre approche sur corpus, le passage entre ces deux dimensions est le fruit d'un travail d'analyse de la part du linguiste/terminologue qui travaille sur corpus, en collaboration avec les experts. Cette analyse ne peut être conçue comme la mise au jour de phénomènes *donnés* par les indices, mais bien comme le résultat de la *construction* d'une interprétation de phénomènes langagiers (par ex. Boutet, 1994 ; 2008 ; Condamines, 2005 ; 2006 ; 2007).

La place de l'interprétation en corpus est de plus en plus admise, notamment par certains travaux de TAL (Habert & Zweigenbaum, 2003) qui remettent ainsi en cause la tradition distributionnelle harrissienne classique (Harris, 1954) qui suppose que le sens émerge des combinaisons de formes en corpus (Condamines, 2003a : 27). Mais plus particulièrement, cette position est tenue par le courant de la terminologie textuelle (Aussenac-Gilles & Bourigault, 2003 ; Aussenac-Gilles, *et al.*, 2002 ; Condamines, 2003a ; 2005 ; 2006 ; 2007) qui insiste sur l'importance de l'interprétation et des paramètres qui la conditionnent. Notamment, Condamines (2003a) met en avant le caractère « doublement situé » de l'interprétation à partir de corpus, à la fois par rapport à la situation de production des textes qui composent les corpus, mais également par rapport aux objectifs de l'étude.

Le premier point renvoie principalement au fait que l'interprétation des phénomènes langagiers extraits est étroitement liée à la nature et au type des corpus sélectionnés. En diachronie, ceci se traduit principalement par le choix des textes (domaine, genre et degré de spécialisation) mais surtout par celui de l'intervalle et de la période des corpus sélectionnés (Chapitre 3). Le second point concerne le fait que l'objectif de l'étude conditionne aussi en

grande part l'interprétation. Ces éléments doivent être pris en compte dans la construction de l'interprétation et sont sous-jacents au dialogue entre les acteurs de la construction (§10.2 *infra*). En effet, l'un des aspects les plus centraux dans l'optique de la construction d'une interprétation pour les tâches terminologiques est que celle-ci doit être le fruit d'un consensus prenant en compte l'objectif applicatif sous-jacent à l'analyse pour guider les acteurs de la construction.

10.1.2 Consensus interprétatif

L'interprétation des données en corpus se fait sur la base d'un jugement et/ou d'une analyse humaine. Dans la mesure où l'analyse est basée sur corpus, il est non seulement clair que l'on ne peut associer un phénomène interne en corpus à un seul phénomène externe dans le domaine, mais également que les connaissances repérables n'y sont que partielles. En effet, malgré tout le soin que l'on peut apporter à la constitution des corpus, ceux-ci ne sont que partiellement représentatifs des connaissances d'un domaine. De plus, les outils utilisés pour les traiter ne permettent de repérer qu'une partie des informations contenues. Il est donc nécessaire d'avoir recours à un analyste humain pour valider les résultats obtenus automatiquement sur corpus, voire pour les compléter. Dans les tâches de constitution de produits terminologiques classiques, en synchronie, il arrive que l'« interprète » des données ne soit qu'une seule personne chargée de valider les résultats obtenus à partir des outils. Or, en diachronie, la construction de l'interprétation ne peut pas ne reposer que sur un expert ou un analyste, mais doit être fondée sur la collaboration de ces deux parties. En effet, et nous montrons pourquoi dans les sections suivantes, aucun des deux acteurs de la construction de l'interprétation des données ne dispose à lui seul de connaissances suffisantes pour parvenir à une analyse fiable et pleinement satisfaisante. Cette analyse doit donc être le fruit d'un consensus entre l'analyste et l'expert. Par ailleurs, ce consensus doit être guidé par les besoins applicatifs de l'analyse, comme nous allons le voir.

Le dialogue entre analyste et expert, l'analyse des corpus ainsi que l'application rentrent dans ce que nous appelons un schéma de co-construction, représenté à la Figure 10.1.

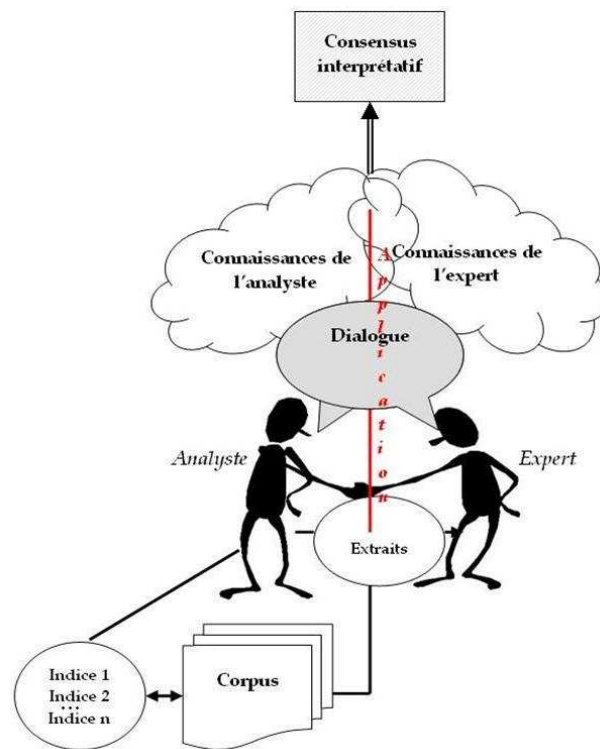


Figure 10.1 – Schéma de co-construction de l'interprétation

Dans ce schéma, le principe de co-construction repose sur :

- un dialogue entre l'analyste et l'expert, chacun disposant de connaissances complémentaires et différentes pour analyser l'évolution d'un domaine ;
- un corpus, auquel l'analyste a accès à travers une analyse linguistique et des indices et auquel l'expert a accès en guise de consultation, *via* des extraits, des contextes et l'analyse préliminaire de l'analyste ;
- une application, transversale, qui guide à la fois la constitution et l'analyse du corpus, éventuellement la sélection des indices, ainsi que le dialogue entre les deux acteurs de la construction, l'objectif final étant de parvenir à un consensus interprétatif sur l'évolution.

Dans les sections suivantes, nous détaillons le rôle et la complémentarité des deux acteurs de la co-construction, ainsi que le rôle de l'application.

10.2 Acteurs de la co-construction

La construction de l'interprétation à partir de corpus diachroniques est guidée par des « acteurs » humains en charge de l'analyse. Dans l'optique choisie dans cette recherche, deux types d'acteurs collaborent : les experts du domaine et l'analyste¹²⁰.

10.2.1 Analyste

10.2.1.1 Définition de l'analyste

Le terme « analyste » renvoie à l'interprète des phénomènes linguistiques repérables en corpus. Celui-ci est généralement un *langagier* (linguiste, terminologue, traducteur, etc.), tel que nous l'avons présenté dans notre recherche, mais peut également être un ingénieur de la connaissance, un ergonome, etc., qui possède une compétence linguistique – c'est-à-dire une compétence de linguiste, d'« expert » des phénomènes linguistiques – et une compétence de l'exploration outillée de corpus.

Ce dernier n'est généralement pas spécialiste du domaine¹²¹ qu'il analyse et ne travaille que sur la base des corpus de textes constitués pour la recherche.

10.2.1.2 Atouts : l'analyse de corpus de textes

Le principal atout de l'analyste est sa compétence linguistique pour l'analyse de corpus et par conséquent, le fait qu'il n'ait aucun *a priori* sur l'évolution du domaine qu'il observe. Plus précisément, ses connaissances sont avant tout des connaissances sur le fonctionnement de la langue, des textes et des phénomènes linguistiques repérables à partir d'indices et d'outils dans les textes et non pas sur le domaine lui-même. Ce manque d'*a priori* lui offre une certaine objectivité sur l'évolution du domaine étudié, objectivité nécessaire que n'a pas l'expert (cf. *infra*).

¹²⁰ Notons qu'un troisième acteur aurait été possible, l'historien des sciences.

¹²¹ À l'exception de certains rares cas tels qu'illustré dans (Bourigault & Lamé, 2002).

De plus, comme le remarque notamment Boutet (1994), la situation de l'analyste présente deux caractéristiques majeures :

- il n'est pas le destinataire des textes qu'il analyse,
- et, à l'inverse d'un formateur qui serait engagé dans une interaction de groupe, l'interprétation pour l'analyste est une activité différée dans le temps, sur une interaction qui a déjà eu lieu.

Ces deux éléments ont pour conséquence le fait que l'interprétation de l'analyste ne peut prendre en compte tous les éléments qui ont présidé à la « naissance » du texte, mais également qu'il ne « lit » pas le texte (Condamines, 2003a : 39). En effet, l'analyste accède au corpus à travers des indices ou des outils, dans une approche qui le « délinéarise ». Il aborde donc les connaissances présentes en corpus sous un angle inédit dont ne dispose pas l'expert. Cette vision du texte lui permet à la fois d'extraire des informations sur l'évolution à partir d'indices linguistiques et de ses connaissances sur le fonctionnement de la langue et des textes, mais également de guider l'expert grâce à ses explorations. En effet, cette approche de découverte de l'évolution à partir de fonctionnements linguistiques lui permet d'amorcer l'exploration et de présenter des pistes de dialogue à l'expert. Il est difficile en effet à ce dernier de cibler des évolutions spécifiques *ex nihilo*, en particulier lorsque l'analyse porte sur un intervalle précis. Nous revenons en détail sur cette question au §10.2.2.3 (p.307).

10.2.1.3 Limites de l'analyste

Mais si l'approche de l'analyste présente des atouts incontestables, elle connaît au moins deux limites. La première est le fait que le corpus construit pour l'analyse n'est nécessairement que partiellement représentatif des connaissances du domaine. La reconstruction que l'analyste en fait à partir d'une étude linguistique est également partielle. Deuxièmement, l'hétérogénéité des interprétations possibles à partir de mêmes indices linguistiques rend nécessaire la discussion avec un expert du domaine dont seules les compétences permettent de sélectionner les informations. La limite alors est le manque de

connaissances de l'analyste sur le domaine étudié. Pour ces différentes raisons, la collaboration avec des experts du domaine s'impose.

10.2.2 Expert du domaine

10.2.2.1 Rôle de l'expert¹²²

La question des experts en terminologie est assez peu abordée dans la littérature et rares sont les définitions proposées sur cette notion. De manière tacite, les études s'accordent généralement pour considérer que l'expert est une « personne dont l'expérience et le haut niveau de compétence sont reconnus dans un domaine particulier¹²³ ». Dans ce sens, Otman (1993 : 76) recense différentes définitions de l'expert pour la terminologie :

« Pour ce qui concerne la terminologie, l'OLF, dans son *Vocabulaire systématique de la terminologie* (R. Boutin-Quesnel *et al.* Québec, 1985) [...] définit le spécialiste comme un « expert d'un métier ou d'une profession associé à des travaux de terminologie à titre de consultant ou d'informateur ». Le spécialiste est un « témoin de l'usage qui permettra d'infirmer ou de confirmer les hypothèses établies par le terminologue à partir de la documentation » (Rondeau, 1981 : 101). Pour D. Gouadec, l'expert est « l'homme de terrain, qui sait de quoi il parle », c'est un « informateur technicien » (1990 : 201). »

Mais définir le rôle de l'expert pour une tâche donnée est une question qui, bien qu'essentielle, est difficile à traiter. Le rôle de l'expert a d'ailleurs grandement évolué au fil du temps et des théories de la terminologie. Ainsi, si les tenants des principes classiques wüstériens placent l'expert au centre du diagnostic et, conformément à la démarche onomosialogique, lui attribuent un rôle essentiel en tant qu'expert « dépositaire » du système conceptuel de son domaine, les approches plus récentes de la socioterminologie et de la terminologie textuelle privilégient en revanche les informations – y compris d'ordre

¹²² Bien que Otman (1993) propose la distinction, nous utiliserons ici de manière indifférenciée les termes « expert » et « spécialiste ».

¹²³ Définition proposée par le Grand Dictionnaire Terminologique, Catégorie « Appellation de personne », Définition 5. Office Québécois de la Langue Française 2007, http://www.granddictionnaire.com/btml/fra/r_motclef/index800_1.asp (consulté le 7 octobre 2008).

conceptuel – issues des textes eux-mêmes et de l’usage en discours. Le rôle que ces théories attribuent à l’expert est donc davantage un rôle de partenaire du linguiste/terminologue (Bourigault & Slodzian, 1999 et Thoiron & Béjoint, 2003 entre autres sur cette évolution).

En synchronie, si ces différents rôles sont en théorie bien balisés en fonction des différents courants, dans les faits les allers-retours constants entre sémasiologie et onomasiologie (Condamines, 1999) rendent nécessairement double le rôle de l’expert : par exemple, celui-ci est « partenaire » lorsqu’il valide les candidats-termes extraits de textes, mais il devient « dépositaire » lorsqu’il doit compléter une liste de termes en fonction de ses connaissances.

En diachronie, l’expert est susceptible de revêtir trois rôles, et ce souvent simultanément :

- un rôle de « dépositaire » de l’histoire de son domaine. Son rôle est alors de livrer des éléments sur son évolution, tels qu’une liste de termes/concepts nouveaux, obsolètes ou ayant évolué ;
- un rôle d’« historien » : en lien étroit avec ce premier rôle, l’expert peut parfois devenir « historien » ou « témoin » de l’histoire de son domaine et raconter, informer l’analyste sur les évolutions qui ont eu lieu dans son domaine, ses théories et ses acteurs ;
- un rôle de « partenaire », lorsqu’il valide la liste de termes évolutifs repérés par l’analyste¹²⁴.

10.2.2.2 Qualités et atouts d’un expert pour la diachronie

Lorsque la question est abordée dans la littérature, les qualités que l’on attribue à l’expert qui devra participer à une recherche donnée sont souvent idéales, comme l’illustre l’extrait suivant :

« Les qualités à rechercher chez l’expert sont : « son autorité intellectuelle indiscutable, complétée par une longue pratique, sa volonté et sa capacité à communiquer sur son savoir, son aptitude à l’introspection, son goût pour la recherche de solutions neuves et la remise en cause permanente de ses acquis, son expérience du travail en groupe, sa disponibilité, sa patience, sa ténacité –

¹²⁴ Ajoutons que parfois l’expert peut également jouer le rôle de « rédacteur » des textes qui composent le corpus.

pour ne pas parler d'endurance – et sa tolérance à l'erreur » (Bonnet, *et al.*, 1986 : 165) ». (Otman, 1993 : 77)

En diachronie, toujours dans une situation idéale, on recherche les mêmes qualités. Mais il en est une plus importante que les autres dans ce contexte : l'ancienneté. En effet, plus un expert est ancien dans son domaine, plus il a d'expérience et plus il peut avoir de recul sur l'évolution de son domaine. L'expert idéal recherché en diachronie doit donc de préférence ne pas être un novice dans son domaine et montrer un goût certain pour l'évolution de celui-ci, évolution dont il peut avoir le sentiment d'être lui-même un acteur.

Dans ces conditions, l'atout majeur de l'expert repose sur ses compétences dans le domaine étudié, mais surtout des connaissances qu'il a de son évolution. Néanmoins, ce dernier aspect est loin d'être facile à maîtriser.

10.2.2.3 Difficultés du « rétro-diagnostic »

En plus des limites pratiques connues lorsque l'on collabore avec des experts du domaine (difficulté à mettre en place des rencontres, manque de temps, manque de disponibilité et parfois manque de considération pour la tâche à laquelle ils contribuent), dans le cas de la diachronie, les difficultés principales du recours aux experts sont liées à la notion de *rétro-diagnostic* (Dury, 2008a).

Le terme « rétro-diagnostic » renvoie au fait que l'expert doit porter un regard sur l'histoire et l'évolution des connaissances dans son domaine. Cela se traduit principalement par le fait qu'il doit solliciter son sentiment d'évolution sur certains concepts et connaissances. Cette notion de *sentiment d'évolution* est à mettre en lien avec celle de *sentiment néologique* abordée par certains néologues. Les membres de l'équipe de Gardin (1974) sont les premiers à la décrire dès 1974, équipe dont l'étude repose sur le recueil des néologismes dans huit pages de l'hebdomadaire *Le Point*. Pour ce faire, les huit pages ont été soumises à des membres de leur équipe avec trois consignes : « 1) soulignez les néologismes ; 2) indiquez les contextes qui vous paraissent nécessaires à leur explicitation ; 3) faites une liste des unités sur lesquelles vous hésitez » (*ibidem* : 45).

D'emblée, les auteurs notent que les consignes 1) et 2) n'ont pas été suivies et qu'elles ne peuvent être prises en compte. Nous retrouvons donc là les difficultés d'élaboration de la tâche mentionnées à la section §9.2.1 (p. 285).

À partir de cette expérience, les auteurs montrent que la compétence de chacun des locuteurs à reconnaître et identifier des néologismes, leur « sentiment néologique », varie non seulement d'un membre de l'équipe à un autre, mais également dans le relevé d'un seul et même membre au cours de sa lecture (*ibidem*, p.46). Ces observations ont été depuis reprises, critiquées et développées notamment par Sablayrolles (2002b ; 2003a) ou Ben Hariz Ouenniche (à paraître)¹²⁵. Sablayrolles (2002b) décrit en particulier certains facteurs qui entrent en jeu dans la variabilité du sentiment néologique (tels que la culture du locuteur, son niveau d'instruction, sa conception d'un néologisme, etc.) et propose de guider et saisir ce sentiment à travers la prise en compte des matrices lexicogéniques : les néologismes proposés à des locuteurs sont répartis en fonction de leurs modes de formation et sont ensuite classés sur une échelle de « néologicit   » en fonction de l'accord entre le sentiment néologique de trois locuteurs, de leur unanimit      leur plus fort d  saccord. L'auteur montre en effet qu'en fonction du mode de formation du n  ologisme, le taux d'accord est plus ou moins grand. L'int  r  t de cette d  marche est que Sablayrolles int  gre ainsi la variabilit   du sentiment n  ologique dans la d  finition m  me de la n  ologie, qu'il d  crit de fait comme ph  nom  ne scalaire. Les diff  rentes analyses que nous avons men  es dans notre travail font   cho    ces   l  ments, ce qui souligne l'importance de mettre en place une m  thodologie qui int  gre plusieurs indices, une analyse linguistique et la collaboration avec des experts.

Comme le sentiment n  ologique, le sentiment d'  volution est un sentiment subjectif qui ne s'appr  cie qu'en fonction des connaissances dont on dispose sur un domaine et son   volution (cf. l'exemple du « Bateau de Th  s  e » (§8.2.4.2)). N  anmoins,    la diff  rence du sentiment n  ologique qui concerne essentiellement la comp  tence « interne » d'un locuteur, le sentiment d'  volution concerne    la fois la comp  tence « externe » des experts et, parfois   galement, leur comp  tence interne.

¹²⁵ Bien que nous ayons   t   en contact avec l'auteur au sujet de ses r  cents travaux, ceux-ci n'ont pu   tre int  gr  s    temps    notre manuscrit.

Cette distinction est encore une fois importante dans la mesure où travailler sur la dimension interne offre l'avantage de présenter un caractère relativement « binaire » pour l'expert : l'unité qu'on lui présente est soit un terme du domaine, soit ce n'est pas le cas¹²⁶. Bien entendu, cette vision du terme est idéale et il est clair qu'en pratique, les réponses fournies par les experts sont rarement « blanches ou noires ». Mais en diachronie, la difficulté est accentuée par le fait que l'objet sur lequel porte le jugement n'est non pas le terme, mais une évolution. En sus de la difficulté à définir nettement ce que l'on appelle « évolution », il est clair que cette notion n'est pas « binaire », mais continue : un objet évolue un peu, beaucoup, suffisamment, rapidement, etc. L'objet qui évolue peut être le terme, le concept, l'instance, une fonction, etc. De plus, comme dans le cas de la néologie, l'évolution est un phénomène transitoire. Et lorsque l'on travaille sur des périodes anciennes, cela implique que l'expert, pour poser un rétro-diagnostic, doit interroger son sentiment sur une évolution qui a pu se produire des années auparavant.

Pour répondre à ces difficultés, une possibilité est de chercher à interroger plusieurs experts pour parvenir à un consensus. En effet, dans le cas de l'histoire d'un domaine, plusieurs points de vue peuvent coexister. Il n'y a alors pas forcément une seule manière de voir les choses, mais une conjonction de points de vue que l'on peut prendre en compte au mieux pour parvenir à un consensus. De fait, dans l'idéal, il est important d'essayer de réunir et d'interroger plusieurs experts afin de confronter et d'harmoniser leurs jugements. À défaut de les réunir¹²⁷ pour « débattre » de leurs avis, il est possible au moins d'interroger plusieurs d'entre eux et l'analyste pourra harmoniser *a posteriori* les données ainsi obtenues. Mais, encore une fois, les experts qui acceptent de participer à ce type de tâches sont rares et l'analyste n'a généralement que peu le choix du nombre d'experts à consulter. De fait, si multiplier les jugements peut être un atout, interroger un seul expert peut déjà s'avérer précieux et riche, et peut parfois être plus facile à gérer pour l'analyste. En effet, un autre point important concerne le fait que les jugements portés sur l'évolution d'un domaine et de

¹²⁶ L'expert ne dira jamais « ceci est un peu un terme ». À défaut, il pourra éventuellement dire plutôt : « ceci n'est pas un terme de mon domaine, mais je l'utilise dans tel et tel contexte ».

¹²⁷ Et il s'agit là d'une réelle difficulté pratique. Dans notre étude notamment, nous n'avons jamais pu réunir les experts.

ses concepts sont souvent instables et difficiles à établir et obtenir un consensus entre plusieurs experts n'est pas toujours aisé (Calberg-Challot, *et al.*, 2007 : 130). Soulignons également que si le consensus entre experts n'est pas toujours possible, ce que l'on pourrait appeler le « consensus individuel », c'est-à-dire les jugements successifs que peut poser un seul et même expert sur un élément donné, est lui aussi instable. En effet, nous avons à plusieurs reprises soumis certains concepts à un seul et même expert et nous avons pu constater que les informations fournies par ce dernier sur leur évolution pouvaient parfois varier. L'analyste doit donc apprendre à traiter des informations hétérogènes d'un expert à l'autre, mais également provenant d'un seul et même expert.

Cette variation s'explique sans doute en grande part à travers la notion d'individualité des experts : chaque expert a sa formation propre, son expérience et ses opinions sur son domaine, qui impliquent nécessairement des regards différents sur la terminologie et les connaissances de ce domaine. Estopà Bajot (2001) souligne clairement ce phénomène en montrant que l'extraction de termes varie très sensiblement en fonction du profil de celui à qui revient la tâche. Dans cette étude, l'expert est considéré comme le professionnel qui, à un moment donné de son activité, fera face à l'extraction de termes, c'est-à-dire non seulement l'expert du domaine, mais également les documentalistes, les traducteurs et les terminologues. Bien que le point de vue de cette recherche soit différent du nôtre, cette comparaison illustre très bien ce phénomène et les résultats que l'auteure présente sont très marqués : le taux d'accord entre groupes varie de 11 à 88% et le nombre de termes extraits peut être multiplié par 4 en fonction du profil du spécialiste qui mène la tâche. De fait, ce type d'études montre que le problème du consensus et de la variabilité du jugement est loin d'être anodin et nécessite que l'analyste interroge des experts de niveau et formation comparables et prenne la décision d'arrêter ce consensus à un moment donné pour mener à bien sa tâche.

Dans notre recherche, lors des différentes tâches proposées, trois types majeurs de désaccords ont été rencontrés à propos de l'évolution des termes/concepts présentés aux experts.

A. Désaccord total

Le premier type de désaccord à distinguer est celui des désaccords totaux. Dans ces cas, les experts ont posé des jugements contradictoires sur un terme/concept donné. Dans notre étude, ce cas s'est posé notamment dans le TTVS pour le terme/concept *Télescope de Korch* qui a été extrait à l'aide de l'indice de fréquence. Le premier des deux experts à qui nous avons montré ce terme (lors de la première série d'entretiens) a jugé ce terme/concept comme néologique, alors que le second expert l'a jugé ancien. Outre le fait que l'analyste devra trancher entre ces deux positions contradictoires ou faire appel à un troisième expert, il est étonnant que de telles oppositions puissent avoir lieu. Dans notre exemple, la raison tient vraisemblablement à une mauvaise connaissance du concept de la part de l'un des deux experts : il a pu donner un jugement aléatoire pour un concept qu'il connaissait peut-être mal, et ce malgré le fait qu'il avait la possibilité de ne pas répondre. Ce type de cas illustre le fait que le sentiment d'évolution peut également être « parasité » par une méconnaissance de certains concepts, en particulier lorsque l'expert manque d'expérience dans le domaine (ce qui a été le cas dans l'exemple rapporté ici).

B. Différentes sensibilités à l'évolution

Le deuxième cas de figure se rattache parfaitement à l'exemple du Bateau de Thésée et renvoie aux différents degrés de sensibilité que peuvent avoir les experts sur l'évolution d'un même terme/concept. Nous prendrons en exemple ici deux cas de figure rencontrés dans le corpus TTVS. Le premier est l'exemple des miroirs. Nous avons proposé le terme/concept *miroir* comme candidat à l'évolution. Deux réactions très différentes ont pu être constatées. La majorité de nos experts ont jugé que les miroirs étaient dans le domaine de l'optique depuis toujours et qu'il n'y avait donc pas d'évolution (bien que certains admettent que leurs propriétés aient changé). Mais, sur le même terme/concept et avec les mêmes indices à l'appui, l'un de nos experts a quant à lui jugé que les miroirs étaient en progrès constants (moins encombrants, plus légers, de meilleure qualité, etc.) et donc qu'il y avait bien une évolution de ce concept. Comme illustré dans les chapitres précédents, nous sommes finalement parvenus à un consensus en mettant en avant le fait que la thématique des miroirs est une thématique rappelée dans le domaine dans la mesure où elle est au cœur de nombreux progrès et évolutions.

Un autre exemple illustre les cas fréquents de concepts liés à des projets spatiaux du moment. Le terme/concept *sondage* par exemple, est un terme lié au projet IASI. Nos experts ont encore une fois été très partagés. Certains ont jugé que ce terme/concept était lié à un projet précis, donc trop ponctuel pour pouvoir parler d'évolution. D'autres à l'inverse ont jugé que ce terme/concept, puisqu'il est lié à un projet central du Cnes, est bien lié à des progrès dans le domaine. Ces deux exemples illustrent les différentes interprétations d'évolution possibles qui peuvent découler de données semblables. Ce cas illustre à la fois la difficulté à saisir l'évolution, la variabilité du sentiment évolutif qui en découle ainsi que sur la nécessité de s'appuyer sur d'autres éléments pour trancher.

C. Hésitations sur les limites de son domaine de compétence

Enfin, le dernier cas de figure est lié à l'évolution du domaine. Comme nous l'avons souligné au Chapitre 8, l'évolution des connaissances d'un domaine ne se résume pas à l'évolution de sa terminologie. Le domaine évolue en tant que tel, et l'optique spatiale en est un parfait exemple à travers sa fusion progressive avec l'électronique. Un nouveau domaine semble donc en train de voir le jour : l'optoélectronique. Cependant, l'apparition d'un nouveau domaine est un processus assez lent, en particulier au niveau de son assimilation par les experts. Ainsi, lorsque nous avons présenté des termes/concepts liés à l'optoélectronique (c'est-à-dire des concepts d'électronique utilisés en optique de manière régulière), quelques experts se sont montrés hésitants. En effet, certains d'entre eux ont jugé que ces termes ne faisaient pas partie de leur domaine de compétence. Ils n'ont donc pas souhaité porter de jugement sur leur évolution. D'autres, au contraire, ont parfaitement assimilé la fusion naissante entre optique et électronique et, bien que tous aient souligné que ce terme faisait aussi partie du domaine électronique, ils n'ont pas hésité à porter un jugement. Dans ce cas, on voit que c'est l'évolution globale du domaine qui entrave le sentiment d'évolution en contraignant les experts à choisir entre un domaine nouveau, quelque peu « hybride », et leur domaine d'origine.

Le rétro-diagnostic est difficile à poser pour un expert du domaine. De fait, le dialogue avec l'analyste permet de guider l'expert dans sa tâche et de lui donner des éléments pour asseoir

son sentiment sur l'évolution. Pour ce faire, il est important de mettre en place les échanges les plus efficaces possible.

10.2.3 Nature des échanges

La mise en place des entretiens avec les experts est une tâche difficile mais incontournable. En effet, si le rôle de l'analyste et l'importance des corpus sont indéniables pour repérer l'évolution en diachronie, il est également essentiel de dialoguer avec les experts pour la compléter. En d'autres termes, il doit y avoir un équilibre entre les données qu'apporte l'analyste à partir de son étude des textes pour guider l'expert, et l'expert qui aiguille le linguiste sur les phénomènes pertinents.

Si les entretiens classiques prennent la forme d'une rencontre entre l'analyste et un ou plusieurs expert(s) (comme illustré au Chapitre 9), il peut arriver que ces entretiens revêtent des formes inédites. Ainsi, en fonction des disponibilités de l'expert, la discussion peut être menée par téléphone sur des questions ponctuelles. Ou encore, en fonction du degré de familiarité entre l'expert et l'analyste, une correspondance par courriel peut s'établir. Mais il est certain que les consignes proposées aux experts changent considérablement en fonction de la nature des entretiens mis en place. Et la qualité des résultats en dépend aussi fortement. La difficulté est donc de maîtriser la consigne donnée à l'expert en fonction du média de communication, ainsi que de « surveiller » la démarche de l'expert. Il est évident que si l'échange par courriel économise du temps à l'expert, l'analyste quant à lui perd des informations précieuses que seul un dialogue peut permettre d'obtenir.

Soulignons enfin que de manière générale, le point de vue selon lequel il est nécessaire d'envisager une démarche collaborative est de plus en plus partagé en TAL, en terminologie et en ingénierie des connaissances. Certaines études mettent d'ailleurs cette proposition au centre de leur approche et intègrent l'interaction avec les experts au cœur du développement de l'application (Hamon & Nazarenko, 2001a ; 2001b ; Ottens, *et al.*, 2007, etc.). La question de l'application doit néanmoins elle aussi être prise en compte pour guider ce dialogue et l'ensemble du repérage des phénomènes d'évolution.

10.3 Prise en compte de l'application

10.3.1 Application comme guide de l'analyse

La nécessité de la prise en compte de l'application pour certaines tâches terminologiques a été soulignée notamment par Aussenac-Gilles, Condamines et Szulman (2002). Les auteures montrent que la nature de l'application en jeu pour une analyse en conditionne chacune des étapes. Leur position est que la construction de produits terminologiques à partir de corpus ne peut pas ne pas être guidée par l'application, sous peine que ces derniers soient construits de manière trop générique et ne soient pertinents pour les utilisateurs. La nature de l'application conditionne entre autres :

- l'analyse des textes : les auteures prennent l'exemple de l'extraction de marqueurs de relation et montrent que certains marqueurs sont plus performants pour repérer la relation de généricité. De fait, ils sont particulièrement pertinents pour des ressources qui s'adressent à des non spécialistes ;
- le choix de la méthode et de la structure de la ressource : à partir d'un article de Aït El Mekki et Nazarenko (2002), les auteurs montrent que pour une application telle que la consultation d'un document électronique la ressource à choisir est un index, dont la structure doit être enrichie de liens sémantiques autres que les relations génériques/spécifiques ou *voir aussi* par rapport à un index classique ;
- le contenu de la structure : à partir de textes, l'interprétation des données peut dépendre de la représentation que se fait chaque individu ; ce qui importe alors est d'expliquer les choix et de les justifier en fonction de l'application : choix des concepts et des relations sélectionnés, place du concept dans la ressource, etc.

En diachronie, la même observation peut être faite : l'application permet de guider l'analyse et le repérage des phénomènes d'évolution. Dans notre recherche, l'objectif d'analyse est large afin de permettre une approche exploratoire, qui est nécessaire à la mise au jour et à l'évaluation des indices, ainsi qu'à la caractérisation des phénomènes d'évolution repérables sur de courts intervalles. Cette perspective a guidé l'ensemble de nos choix, que ce soit les

corpus (période, intervalles, lien avec les projets spatiaux), la sélection des candidats-termes, les entretiens avec les experts, la typologie des phénomènes proposés, etc. présentés tout au long de ce travail.

Néanmoins, comme suggéré dans notre étude, de nombreuses applications précises sont directement concernées par l'apport de la perspective diachronique et du repérage de l'évolution des connaissances, telles que la mise à jour de ressources terminographiques, la veille scientifique et technique, la traduction, la didactique des langues de spécialité, etc. Chacune de ces applications peut tirer profit de la démarche mise en place dans cette étude, mais chacune conditionne de manière différente les éléments méthodologiques proposés, dont :

- le choix des périodes et intervalles : si le suivi des évolutions dans le cas de projets spatiaux impose des intervalles courts, l'aide à la traduction peut privilégier des périodes plus anciennes et des intervalles plus longs en fonction des besoins de traduction ;
- le choix des phénomènes d'évolution à proposer : la veille scientifique et technique sera plus intéressée par les évolutions liées à la nouveauté et notamment à l'apparition de concepts ou l'aide à la rédaction bénéficiera plus d'informations liées à la forme des termes ;
- le choix des indices à mettre en œuvre : puisque la forme des termes est l'information privilégiée pour l'aide à la rédaction, les contextes riches en connaissances évolutives sont moins pertinents que la fréquence ou les variantes. De plus, bien qu'à ce stade nous n'ayons pu mettre au jour des combinaisons d'indices spécifiques pour une évolution donnée, cette perspective n'est pas à exclure et permettra de choisir les combinaisons et les indices précis en fonction des besoins de repérage d'une application donnée.

En guise de perspective, nous illustrons dans la section suivante les choix méthodologiques qui peuvent être faits à partir de notre proposition méthodologique, à travers l'exemple de la mise à jour de ressources termino-ontologiques.

10.3.2 L'exemple de la mise à jour de ressources termino-ontologiques

Le terme « ressource termino-ontologique » (RTO) est un terme générique qui regroupe un ensemble de ressources allant

« des thésaurus et langages documentaires utilisés en sciences de l'information aux ontologies manipulées par l'ingénierie des connaissances, en passant par les terminologies et autres bases de données terminologiques ou sémantiques servant à l'étude des langues, au traitement informatique de l'écrit ou encore aidant à la traduction. » (Aussenac-Gilles, *et al.*, 2007b : 7)

Parmi ces différents types de ressources, ces dernières années ce sont les travaux portant sur la question des ontologies qui prêtent une attention nouvelle et toute particulière à la question de leur maintenance dans le temps. Dès 1994 avec les travaux de Cimino *et al.* (Cimino, 1996 ; Cimino, *et al.*, 1994) ou ceux de Oliver *et al.* (1999), certaines recherches soulèvent la question de la gestion des ontologies et proposent principalement des typologies des changements susceptibles d'affecter différentes ontologies (médicales essentiellement). Depuis, d'autres travaux ont vu le jour, qui, en plus de caractériser les changements de l'ontologie, proposent la mise en place effective de différentes méthodes pour maintenir ces ontologies à jour et/ou en gérer différentes versions (par ex. Klein, 2004 ; Luong, 2007 ; Rogozan, 2008 ; Stojanovic, 2004). Ainsi, les différents auteurs s'accordent sur une typologie des phénomènes d'évolution à prendre en compte pour la maintenance des ontologies tels que (Oliver, *et al.*, 1999) :

- l'ajout d'un concept,
- la suppression d'un concept,
- le changement de place d'un concept dans la hiérarchie,
- le changement de relation entre deux concepts,
- et la fusion de deux concepts.

Cependant, comme le remarque notamment Tissaoui (2009), ces travaux ne prennent en compte que l'ontologie et mettent de côté la terminologie. Ils sont en effet très orientés sur la formalisation des ontologies et le côté strictement informatique de leur gestion et se

concentrent sur la composante conceptuelle des ontologies, au détriment de la composante terminologique. Or, l'intérêt de la notion de « RTO » est justement de mettre en avant la nécessité d'associer les composantes conceptuelles et terminologiques dans une ressource afin de garantir des possibilités d'annotation et d'indexation des documents¹²⁸ (entre autres Aussenac-Gilles & Bourigault, 2003 ; Bourigault & Aussenac-Gilles, 2003 ; Bourigault, *et al.*, 2004 ; Buitelaar, *et al.*, 2005 ; Gillam, *et al.*, 2005 ; Maedche & Staab, 2001).

Si la prise en compte des textes pour la construction des RTO est de mieux en mieux balisée, elle reste une question récente et encore peu traitée pour leur maintenance et leur gestion (Aussenac-Gilles, *et al.*, 2007a). C'est ce que déplore Tissaoui (*ibidem*) qui, dans sa thèse en cours, propose entre autres d'adapter la typologie des changements de Stojanovic (*ibidem*) pour prendre en compte la composante terminologique des RTO (dont les changements de dénomination pour un concept donné).

Une des propositions récentes pour intégrer l'analyse des textes dans le cadre de la maintenance des RTO est celle de Gargouri (2003 ; 2009 ; 2004) qui développe la mise en place d'une « passerelle » entre les textes et une ontologie courante. Plus précisément, à partir d'une analyse de textes nouveaux, il propose un modèle pour identifier les concepts nouveaux et les relations nouvelles entre eux, puis pour automatiser leur intégration dans l'ontologie. Cette démarche est appelée par l'auteur « Raffinement Conceptuel par Analyse Vectorielle ». Ce modèle dépasse le cadre de cette recherche, mais nous le mentionnons ici parce qu'il trouve un double écho par rapport à notre propre perspective :

- d'abord parce que l'auteur inscrit la maintenance de ressources dans une démarche d'analyse textuelle,
- ensuite parce que le modèle de Gargouri est semi-automatique et vise à assister les experts du domaine dans la tâche de construction et de maintenance des ontologies.

Ce point de vue met donc en avant à la fois l'intérêt des textes pour la construction et la maintenance de ressources et la nécessité d'une démarche collaborative.

Ces arguments se retrouvent dans une autre proposition, celle d'Ottens *et al.* (2007), qui proposent la conception de Dynamo, un outil de collaboration pour la construction et la

¹²⁸ Ou encore, pour le dire autrement : « partir des textes pour mieux y retourner » (<http://www.unice.fr/urfist/IMG/ppt/WebSemantique3-2006.ppt>, école TCAN, Bourigault et Aussenac-Gilles, 2006, consulté le 15 juin 2009).

gestion d'ontologies. Les auteurs exploitent les progrès de la construction d'ontologies à partir de textes à l'aide d'outils de Traitement Automatique des Langues, dont Syntex utilisé pour notre recherche. L'intérêt de cette approche par rapport à notre perspective est triple :

- premièrement, il met les textes et l'outillage TAL au centre du processus,
- ensuite, il met en avant un processus de construction et de maintenance semi-automatique et collaboratif pour assister les experts,
- enfin, il propose de considérer la maintenance des ontologies en continuité avec leur construction.

Nous retrouvons donc dans ces deux premiers points (relevés également dans les travaux de Gargouri) deux des principes centraux mis en avant dans notre approche et qui permettent de souligner l'apport que celle-ci peut représenter pour la mise à jour des RTO. Les éléments méthodologiques que nous proposons dans notre étude s'inscrivent de fait parfaitement dans ce type de cadre où les textes sont au centre de l'application. De la même manière que la construction de RTO à partir de textes est à l'origine du dialogue entre ingénierie des connaissances et terminologie textuelle, la question de la mise à jour de ces ressources permet bien de poser les jalons de la prise en compte de la dimension diachronique en terminologie textuelle.

Plus précisément, à la lecture des besoins liés à la mise à jour de RTO, notre proposition linguistique peut apporter des éléments méthodologiques sur plusieurs points :

- les changements à cibler,
- la manière de cibler les changements à partir de textes,
- le moment où prendre en compte l'analyse des textes.

Comme nous l'avons vu *supra*, les changements susceptibles d'intervenir dans les RTO sont de mieux en mieux balisés. Ainsi, il est plus facile de se concentrer sur le repérage précis de ces éléments et d'adapter les indices et la construction des corpus d'étude en fonction des besoins. En effet, une fois les types de changements circonscrits, il est plus facile de mener la recherche et notamment de guider les experts mobilisés sur la tâche.

Mais l'impact le plus important du type d'application pour le repérage de l'évolution concerne le choix du moment où effectuer le repérage. En effet, dans les travaux mentionnés *supra*, la gestion des évolutions intervient essentiellement sur une ontologie existante dont on suppose qu'elle est obsolète. Or, si ce point de vue est inévitable, dans la mesure où la maintenance et la construction d'une RTO peuvent être considérées dans une relative continuité (Ottens, *et al.*, 2007), la maintenance de la ressource peut être anticipée au moment de la construction. Par exemple, les contextes riches en connaissances d'évolution pourraient être projetés en corpus dès le moment de la construction de la ressource. Ainsi, les termes/concepts marqués par une évolution au moment de la construction pourraient faire l'objet d'un traitement particulier ou, à défaut, d'un étiquetage permettant d'en assurer le suivi de manière privilégiée pour adapter la ressource dans le temps. De la même manière, la présence de variantes terminologiques pourrait aussi être archivée dans une « fiche » au moment de la construction de la ressource et permettrait ainsi de marquer comme « potentiellement évolutif » un terme/concept concerné par de nombreuses variantes.

Tous ces éléments ne sont que des pistes méthodologiques, mais ils prouvent l'impact qu'une approche telle que la nôtre peut revêtir d'un point de vue applicatif. Les éléments proposés dans cette étude ouvrent également des pistes intéressantes pour répondre à des besoins plus spécifiques qui guident en retour les aspects méthodologiques à privilégier pour repérer l'évolution. De manière générale, cette démarche initie un nouveau champ d'investigation en terminologie textuelle, celui de la dimension diachronique, où beaucoup de pistes restent à explorer.

Conclusion générale

Dans cette thèse, nous avons présenté la mise en place d'une méthode linguistique pour repérer l'évolution des connaissances spécialisées. Ce point de vue a permis de nourrir une réflexion rarement posée sur la place de la perspective diachronique en langue de spécialité. Pour bien comprendre l'enjeu de cette recherche, nous avons en premier lieu dressé un portrait croisé des études diachroniques en langue générale et en langue de spécialité. Cette première étape était nécessaire pour souligner les bouleversements théoriques propres à la terminologie et qui permettent depuis peu d'envisager une réelle réflexion sur la diachronie en langue de spécialité. Sur la base de ce constat, nous avons choisi d'inscrire notre étude dans la lignée de la terminologie textuelle. Nous avons fondé notre recherche sur l'hypothèse forte d'un parallèle entre évolution dans la langue et évolution des connaissances, c'est-à-dire sur l'idée que des évolutions linguistiques repérables en corpus peuvent révéler des évolutions de connaissances d'un domaine. Après avoir détaillé la méthodologie de constitution et d'exploration des corpus mise en œuvre pour tester la validité de cette hypothèse, quatre indices linguistiques associables à des évolutions de connaissances ont été identifiés, définis et analysés. À l'aide de ces indices, nous avons présenté et caractérisé différents phénomènes d'évolution repérables en diachronie courte dans nos corpus et ainsi validé la pertinence de notre hypothèse et des indices mis au jour. Enfin, nous avons clos notre recherche en montrant que l'analyse linguistique des corpus doit être envisagée comme un processus de co-construction, c'est-à-dire de la construction conjointe de l'interprétation des évolutions en corpus de la part de l'analyste et d'experts de domaine.

L'étude que nous avons proposée constitue une contribution majeure sur la question de la diachronie dans les langues de spécialité. Les apports de notre travail peuvent se décliner en trois points, applicatifs, descriptifs et théoriques.

D'un point de vue appliqué, notre principale contribution concerne l'élaboration d'une méthode linguistique outillée en corpus pour repérer l'évolution. L'un des points forts de la méthode proposée est qu'elle s'appuie sur l'utilisation d'outils classiques d'exploration de corpus, tels qu'un concordancier et un extracteur de termes/de relations. Ces outils sont répandus, faciles d'utilisation et souvent mis à contribution dans les tâches classiques d'analyse terminologique. Ce choix méthodologique inscrit notre étude dans la lignée de ces travaux et garantit la reproductibilité de notre démarche, ainsi qu'un faible coût de mise en œuvre (tant du point de vue des outils que de la formation de l'analyste qui mène la tâche de repérage).

Notre méthode repose sur l'exploration de quatre indices linguistiques associables à des évolutions de connaissances :

- l'observation d'empreintes de fréquence, c'est-à-dire de certaines variations de fréquence des termes dans le temps,
- l'extraction de contextes riches en connaissances évolutives, c'est-à-dire des portions de textes qui contiennent des informations pertinentes concernant l'évolution des connaissances,
- l'analyse de variantes terminologiques, c'est-à-dire la coexistence ou la succession de dénominations équivalentes pour un même concept,
- et les dépendances syntaxiques, c'est-à-dire l'observation des changements dans la cooccurrence syntaxique de certains termes dans le temps.

Dans les chapitres 4 à 7, la pertinence de chacun de ces indices pour repérer l'évolution a été démontrée à l'aide d'analyses détaillées de l'évolution dans nos corpus, confirmant ainsi notre hypothèse d'un parallèle entre évolution dans la langue et évolution des connaissances. Il est apparu clairement en effet que chacun de ces indices peut être associé de manière régulière et significative à des évolutions de connaissances. Ils sont cependant très différents et peuvent intervenir à des moments distincts de l'analyse ; certains sont en effet apparus plus appropriés pour débiter le repérage alors que d'autres permettent de cibler des aspects d'évolution très précis à ne prendre en compte que dans un second temps. Par exemple, l'analyse des formes chrono-homogènes et des contextes riches en connaissances évolutives permet d'adopter un point de vue global sur les données, ce qui

constitue un point d'entrée intéressant pour commencer l'exploration de l'évolution et « prendre contact » avec les textes et le domaine. Cette approche globale peut ensuite être affinée par l'observation d'indices locaux tels que la fréquence de termes isolés ou les variantes.

Cet aspect montre l'importance de chercher à combiner les indices pour mieux cerner l'évolution. Ceci constitue l'une des observations saillantes dans notre recherche : en effet, si chaque indice permet bien de mettre au jour de nombreux aspects d'évolution, les combiner permet d'améliorer considérablement les possibilités de repérage. Nous avons observé en effet que, pour un indice linguistique, plusieurs interprétations d'évolution sont envisageables simultanément. De fait, combiner les indices en faisceaux permet de disposer d'une plus grande masse d'informations sur une évolution donnée afin de *décider* de sa pertinence et de la caractériser le plus finement possible.

L'idée d'une « prise de décision » à partir de l'observation d'indices est fondamentale et a permis d'argumenter sur la nécessité d'envisager le repérage de l'évolution en corpus comme un processus de *co-construction*. Cette position n'est pas triviale et nous a amenée à affirmer que le parallèle existant entre langue et connaissances n'est pas « donné » et unique dans les textes, mais doit être construit par l'analyste à partir des indices linguistiques et en collaboration avec des experts de domaine. L'un des apports clés de notre réflexion se situe donc dans la mise en évidence de la pertinence d'une analyse de corpus pour repérer l'évolution, sous réserve que celle-ci assume l'importance de l'interprétation dans l'analyse. Elle est en effet constitutive d'une approche outillée sur corpus dans la mesure où les outils, aussi performants soient-ils, n'identifient pas tous les phénomènes et, surtout, ne permettent pas de les expliquer. De la même manière, les corpus sont toujours une représentation partielle des connaissances. Pour ces différentes raisons, il est essentiel de ne pas « nier », ni minimiser cette étape d'interprétation. Il ne faut pas non plus l'envisager comme une conséquence négative de l'approche sur corpus, mais plutôt la considérer comme l'un des points forts de l'analyse linguistique : la co-construction de l'interprétation met en avant un équilibre entre analyste, experts, indices, corpus et application, garant d'une analyse qui réponde efficacement aux besoins applicatifs définis pour une recherche donnée.

D'un point de vue descriptif, nous avons souligné dans la première partie de notre étude le manque considérable de descriptions et de données sur la nature de l'évolution des connaissances et des langues de spécialité. Une contribution essentielle de notre travail réside dans la typologie établie des phénomènes d'évolution repérés dans nos corpus, qui offre des éléments descriptifs inédits en langue de spécialité. Notre description est particulièrement originale dans la mesure où nous nous sommes intéressée à l'évolution des connaissances en *diachronie courte*, c'est-à-dire à l'évolution en jeu dans de très courts intervalles temporels. Bien que ces derniers ne soient que rarement étudiés, leur pertinence est indéniable pour observer l'évolution des langues de spécialité. Notamment, l'évolution sur de courts intervalles est loin d'être marginale et s'est manifestée à travers 16 types d'évolutions repérables dans nos corpus. D'un point de vue qualitatif, la diversité de la nature de ces évolutions doit être soulignée. Elle intervient à des niveaux multiples, tant du point de vue linguistique (évolution graphique, morphologique, lexicométrique, etc.), que du point de vue extralinguistique (évolution du concept, de l'instance, de la documentation, etc.). Cette diversité confirme l'importance des phénomènes d'évolution en langue de spécialité et la nécessité de mettre en place des descriptions.

Enfin, un apport significatif de cette étude est notre contribution à la réflexion théorique sur la prise en compte de la dimension diachronique en langue de spécialité. Nous l'avons souligné dans la première partie de ce manuscrit, la diachronie est aujourd'hui encore le parent pauvre de la terminologie. Nous nous sommes efforcée de montrer les lacunes que cette situation entraîne pour la connaissance du fonctionnement des langues de spécialité et la nécessité de rendre à la diachronie la place qui lui est due. Notre argumentation s'est appuyée sur le constat du renouvellement théorique que la terminologie connaît depuis une vingtaine d'années et qui offre toutes les bases nécessaires pour développer cette question. Nous avons choisi d'inscrire notre démarche dans le champ de la terminologie textuelle, dans laquelle la dimension diachronique n'a encore jamais été développée. Ce choix nous a amenée à questionner en particulier le lien existant entre textes et connaissances. Pour cela, nous avons montré que les langues de spécialité se démarquent de la langue générale en fondant leur objet d'étude sur une association forte entre dimension interne (linguistique) et

dimension externe (extralinguistique), association que la terminologie textuelle exploite précisément pour accéder aux connaissances à partir des textes. La même hypothèse a été validée en diachronie en montrant que l'évolution linguistique repérable dans les textes est effectivement associable à l'évolution des connaissances. Cette hypothèse a permis de nourrir la réflexion quant à la nature du lien existant entre langue et connaissances. En particulier, nous avons montré l'importance de ne pas restreindre la dimension interne au terme et la dimension externe au concept, mais de l'élargir plutôt au terme pris dans le système de la langue et à la connaissance dans toute sa diversité. Cette affirmation doit être mise en perspective avec le fait que le lien entre dimension interne et dimension externe n'est pas unique et stable mais est le fruit d'une construction de la part de l'analyste et de l'expert. Ce point de vue permet d'aborder une autre question primordiale en terminologie mais également peu traitée : celle de la place de l'expert dans les tâches d'analyse terminologique. À la lueur de la perspective diachronique, cette étude a tenté d'illustrer et de redéfinir le rôle de l'expert et a montré l'importance de ne pas faire reposer l'analyse uniquement sur la connaissance des experts. Il faut les considérer toutefois pleinement comme co-acteurs de l'analyse.

Notre étude ouvre de nombreuses pistes de recherche à explorer. Parmi ces pistes, la première concerne la revalidation de la méthode mise en place. En effet, il est primordial de tester notre démarche sur de nouveaux corpus appartenant à des domaines, genres et périodisations différentes. Ces nouvelles analyses permettraient également de mieux cerner l'influence de ces facteurs pour interpréter l'évolution et en particulier l'influence de la nature du domaine quant à la « relativité » de la nouveauté. D'un point de vue descriptif, l'application de notre méthode sur d'autres corpus ouvrirait également la voie à la mise au jour de nouveaux phénomènes, qui viendraient compléter la première typologie de l'évolution proposée dans notre étude.

De la même manière, chaque indice proposé peut faire l'objet d'une description et d'une mise en œuvre beaucoup plus poussées. Notamment, de nouveaux marqueurs peuvent être

définis pour repérer des contextes riches en connaissances évolutives et pourraient intégrer la dimension discursive dans l'analyse. Une autre piste possible concerne les cas où l'exploration sera menée sur de plus gros corpus et où les dépendances syntaxiques pourraient donner lieu à une analyse distributionnelle pour l'extraction de classes sémantiques telles que Habert & Zweigenbaum (2003) les ont par exemple décrites. On pourrait alors comparer l'évolution de ces classes en diachronie.

Dans le même ordre d'idées, il est envisageable de découvrir d'autres indices linguistiques associables à une évolution des connaissances. En ce sens, une perspective riche serait notamment de développer des indices pour repérer le changement sémantique, qui reste difficile à traiter en corpus. Une piste que nous souhaiterions explorer consiste en l'identification de mécanismes du changement sémantique à l'aide d'une annotation en corpus des actants telle que proposée par Bae *et al.* (2007). Ce type d'annotation se fonde sur les principes de l'approche lexico-sémantique (L'Homme, 2004a) qui s'inspire des Fonctions Lexicales de la Lexicologie Explicative et Combinatoire (Mel'čuk *et al.*) pour la description du sens en langue de spécialité. L'annotation des actants dans un corpus diachronique pourrait permettre d'y observer les évolutions sémantiques, de les décrire et de modéliser certains mécanismes en jeu lors de l'évolution du sens des termes (changement d'actants, changement de rôle pour un actant donné, etc.).

D'un point de vue méthodologique, on peut chercher également à travailler à partir de la comparaison d'autres types de ressources terminologiques. Par exemple, certaines bases de données terminologiques conservent une trace des modifications apportées aux fiches terminologiques. On pourrait alors comparer les données déplacées ou supprimées des fiches et voir les justifications proposées par les terminologues.

Il faudrait également chercher à préciser l'influence de l'application dans le processus de repérage. Nous en avons ébauché un exemple à travers la mise à jour de ressources terminologiques et il est clair que le type d'application en jeu a une influence, tant sur le choix des indices que sur les évolutions recherchées, ou encore sur les interprétations à construire.

Nous avons constaté également que de plus en plus d'applications mettent en avant leur besoin de recherche diachronique. Existe-t-il alors autant de points de vue sur l'évolution que d'applications, ou peut-on parvenir à un certain niveau de « généralité » dans les analyses?

Enfin, l'une des voies de recherche les plus importantes reste de continuer à nourrir la réflexion théorique en diachronie dans les langues de spécialité. Nous espérons avoir convaincu ici de la pertinence et de la nécessité de mettre en place des études diachroniques en terminologie et avons tenté de « mettre à plat » plusieurs questions centrales pour fonder les bases d'une approche diachronique. Mais cette question reste un vaste champ d'exploration, crucial pour montrer l'intérêt et la spécificité de la diachronie en langue de spécialité.

Références bibliographiques

- ACHARD, P., GRUENAI, M. P. & JAUHIN, D. (Éd.) (1985) *Histoire et linguistique*. Édition de la Maison des Sciences de L'Homme, Paris.
- ACOT, P. (1999) *L'histoire des sciences*. Que sais-je?, Presses Universitaires de France (PUF), Paris.
- AHMAD, K. (1996) « A Terminology Dynamic and the Growth of Knowledge: A Case Study in Nuclear Physics and in the Philosophy of Science ». In *Actes de la conférence internationale "Terminology and Knowledge Engineering" (TKE 1996)*, Vienne, Autriche, 26-28 août 1996, pp.1-11.
- AHMAD, K. & AL-THUBAITY, A. (2003) « Can Text Analysis Tell us Something about Technology Progress? ». In *Actes de la conférence internationale "Association for Computational Linguistics" (ACL 2003)*, atelier "Patent Corpus Processing", Sapporo, Japon, 7-12 juillet 2003, pp.46-55.
- AHMAD, K. & MUSACCHIO, M. T. (2004) « Discovery of (New) Knowledge and the Analysis of Text Corpora ». In *Actes de la 4^{ème} conférence internationale « Language Resources and Evaluation »*, Lisbonne, Portugal, 24-30 mai 2004, pp.1567-1570.
- AHMAD, K., SCHIERZ, A. & AL-THUBAITY, A. (2002) « Discovery and Terminology ». In *Actes de la conférence internationale "Terminology and Knowledge Engineering" (TKE 2002)*, Nancy, France, 28-30 août 2002, pp.1-6.
- AÏT EL MEKKI, T. & NAZARENKO, A. (2002) « Comment aider un auteur à construire l'index d'ouvrage ? ». In *Actes du Colloque International sur la Fouille de Textes (CIFT 2002)*, Hammamet, Tunisie, octobre 2002, pp.141-157.
- ALBER-DEWOLF, R. (1982) *Étude sur la création néonymique. Analyse comparée des procédés morphologiques et morphosyntaxiques de formation des termes du domaine de la spectroscopie en anglais, en français, en allemand et en russe*. Thèse de doctorat en linguistique (faculté de lettres), Laval, Québec, Canada.
- ANTHONY, L. (2005) « AntConc: Design and Development of a Freeware Corpus Analysis Toolkit for the Technical Writing Classroom ». In *Actes de la conférence "Professional Communication Conference" (IPCC 2005)*, Limerick, Irlande, 13 juillet 2005, pp.729-737.

- ANTOINE, G. (1981) « L'histoire de la langue. Problèmes et méthodes ». *Le Français Moderne*, 49(2), pp.145-160.
- AUDIBERT, L. (2003) « Étude des critères de désambiguïsation sémantique automatique : résultats sur les cooccurrences ». In *Actes de la 10^{ème} conférence "Traitement Automatique des Langues Naturelles" (TALN 2003)*, Batz-sur-Mer, France, 11-14 juin 2003, pp.35-44.
- AUGER, P. (1973) *Le vocabulaire forestier au Québec, des origines à nos jours. Étude historique*. Thèse de doctorat en linguistique, Université de Strasbourg.
- AUGER, P. (1999) *L'implantation des officialismes halieutiques au Québec, essai de terminométrie*. Langues et sociétés, Office Québécois de la Langue Française (OQLF), Québec, Canada.
- AUROUX, S. (1998) *La raison, le langage et les normes*. Presses Universitaires de France, Paris.
- AUSSENAC-GILLES, N. (2004) « Représentation sémantisée des textes : terminologies et dimensions pragmatiques (qui-quand-où) ». In *Actes du Colloque International sur la Fouille de Textes (CIFT 2004)*, La Rochelle, France, 22-24 juin 2004, pp.33-39.
- AUSSENAC-GILLES, N. & BOURIGAULT, D. (2003) « Construction d'ontologies à partir de textes ». In *Actes de la 10^{ème} conférence "Traitement Automatique des Langues Naturelles" (TALN 2003)*, Batz-sur-Mer, France, 11-14 juin 2003, pp.27-50.
- AUSSENAC-GILLES, N., CONDAMINES, A. & SÉDÈS, F. (Éd.) (2007a) *Évolution et maintenance des ressources termino-ontologiques*. Information – Interaction – Intelligence (I3), Hors série.
- AUSSENAC-GILLES, N., CONDAMINES, A. & SÉDÈS, F. (2007b) « Évolution et maintenance des ressources termino-ontologiques : une question à approfondir ». In N. Aussenac-Gilles, A. Condamines & F. Sédès (Éd.), *Ressources termino-ontologiques*, Information – Interaction – Intelligence (I3), Hors série, pp.7-14.
- AUSSENAC-GILLES, N., CONDAMINES, A. & SZULMAN, S. (2002) « Prise en compte de l'application dans la constitution de produits terminologiques ». In *Actes des 2^{èmes} assises nationales du groupe de recherche I3 (Information, Interaction, Intelligence)*, J. Le Maître (Éd.), Cepaduès, Nancy, France, pp.289-303.
- AUSSENAC-GILLES, N. & SÉGUÉLA, P. (2000) « Les relations sémantiques : du linguistique au formel ». *Cahiers de Grammaire, « Sémantique et Corpus »*, 25, pp.175-198.

- BACHELARD, G. (1938) *La formation de l'esprit scientifique*. Vrin, Paris.
- BACHIMONT, B. (2000) « Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en ingénierie des connaissances ». In J. Charlet, M. Zacklad, G. Kassel & D. Bourigault (Éd.), *Ingénierie des connaissances, évolutions récentes et nouveaux défis*, Eyrolles, Paris, pp.305-324.
- BAE, H. S., L'HOMME, M.C. & LAPALME, G. (2008) « Semantic role in multilingual terminological descriptions: Application to French and Korean Contexts ». In *Actes de l'atelier "Multilingual and Comparative Perspectives in Specialized Language"*, conférence internationale "Language Resources and Evaluation" (LREC 2008), Marrakech, Maroc, 26 mai-1 juin 2008.
- BANEYX, A., MALAISE, V., CHARLET, J., ZWEIGENBAUM, P. & BACHIMONT, B. (2005) « Synergie entre analyse distributionnelle et patrons lexico-syntaxiques pour la construction d'ontologies différentielles ». In *Actes de la conférence internationale "Terminologie et Intelligence Artificielle" (TIA 2005)*, Rouen, France, 4-5 avril 2005.
- BANKS, D. (2003) « L'évolution de la phrase en anglais scientifique ». In *Actes des 4^{èmes} Journées de l'Équipe de Recherche en Linguistique Appliquée (ERLA 2003)*, Brest, France, 12-14 novembre 2003, pp.203-221.
- BARONI, M. & EVERT, S. (2007) « You shall know a word by the company it keeps! Collocation extraction with statistical association measures ». Cours d'introduction donné lors de l'école d'automne "DGfS Fall School in Computational Linguistics", 3-14 septembre 2007, Potsdam, Allemagne.
- BARRIÈRE, C. (2001) « Investigating the Causal Relation in Informative Texts ». *Terminology*, 7(2), pp.135-154.
- BARRIÈRE, C. & AGBAGO, A. (2006) « TerminoWeb: A Software Environment for Term Study in Rich Contexts ». In *Actes de la conférence internationale "Terminology, Standardisation and Technology Transfer" (TSTT 2006)*, Beijing, Chine, 25-26 août 2006, pp.103-113.
- BAUDET, J.-C. (1989) « Histoire du vocabulaire de spécialité, outil de travail pour l'historien des sciences et des techniques ». In *Actes du colloque "Terminologie diachronique"*, C. De Schaetzen, Bruxelles, Belgique, 25-26 mars 1988, pp.56-67.
- BEACCO, J. C. & MOIRAND, S. (1995) *Les enjeux des discours spécialisés*. Presses de la Sorbonne Nouvelle, Paris.

- BÉCIRI, H. (2000) « Néologie générale, néologie spécialisée : quelques observations dans le domaine de la micro-informatique ». *La Banque des mots*, 60, pp.19-36.
- BÉJOINT, H. (1989) « À propos de la monosémie en terminologie ». *Meta, Presses de l'Université de Montréal*, XXXIV(3), pp.405-411.
- BÉJOINT, H. (2004) « Compte rendu : Gaudin, F. (2003) : Socioterminologie, une approche sociolinguistique de la terminologie, Bruxelles, Duculot De Boeck, 286 p. ». *Meta*, XLIX(2), pp.384-387.
- BELTRAN-VIDAL, D. & MANIEZ, F. (Éd.) (2005) *Les mots de la santé*. Presses Universitaires de Lyon, Travaux du CRTT (Centre de Recherche en Terminologie et Traduction), Lyon.
- BEN HARIZ OUENNICHE, S. (à paraître) « Sentiment néologique : protocole et résultats ». *Neologica, Actes de la Journée Conscla du 16 mars 2007*, 3.
- BENVÉNISTE, É. (1966) *Problèmes de linguistique générale*. Bibliothèque des Sciences Humaines, Gallimard, Paris.
- BERTELS, A. (2005) « Les spécificités en contexte : comment étudier la polysémie dans un corpus technique? ». In *Actes des 7^{èmes} journées scientifiques du Réseau "Lexicologie, Terminologie, Traduction" (Agence Universitaire de la Francophonie) : « Mots, termes et contextes »*, Bruxelles, Belgique, 8-10 septembre 2005, pp.371-380.
- BERTELS, A. (2006) *Polysemy in a Technical Lexicon. A Quantitative Study*. Thèse de doctorat en linguistique, Leuven, Belgique.
- BESSÉ (DE), B. (2000) « Le domaine ». In H. Béjoint & P. Thoiron (Éd.), *Le sens en terminologie*, Presses Universitaires de Lyon, Travaux du CRTT (Centre de Recherche en Terminologie et Traduction), Lyon, pp.182-197.
- BIBER, D. (1993) « Representativeness in Corpus Design ». *Literary and Linguistic Computing, Oxford University Press*, 7(4), pp.243-257.
- BLACKWELL, S. (1993) « From Dirty Data to Clean Language ». In B. Aarts, P. Haan (de) & N. Oostdijk (Éd.), *English Language Corpora: Design, Analysis, Exploitation*, Rodopi, Amsterdam, pp.97-106.

- BLANK, A. (2001) « Pour une approche cognitive du changement sémantique lexical : aspect sémasiologique ». In J. François (Éd.), *Théories contemporaines du changement sémantique*, Peeters, Leuven.
- BLANK, A. (2003) « Words and Concepts in Time: Towards Diachronic Cognitive Onomasiology ». In R. Eckardt, K. VonHeusinger & C. Schwarze (Éd.), *Words in Time : Diachronic Semantics from Different Points of View*, Mouton de Gruyter, Trends in Linguistics, Studies and Monographs, 143, Berlin, New York.
- BLUMENTHAL, P. (2008) « Histoires de mots : affinités (s)électives ». In *Actes du 1^{er} Congrès Mondial de Linguistique Française (CMLF 2008)*, Paris, France, 9-12 juillet 2008, pp.31-46.
- BONHOMME, M. (1987) *Linguistique de la métonymie*. Peter Lang, Francfort/Berlin/Bern/New York/Paris/Vienne.
- BONNAL-VERGES, F. (1999) *Quand faire, c'est dire... Essai sur la néologie en LSF en 1999*. Mémoire de Maîtrise, Université Toulouse 2.
- BONNAL-VERGES, F. (2006) « Étymologie et langue des signes française (LSF) ». *Bulag (Bulletin de Linguistique Appliquée et Générale)*, 31, pp.149-164.
- BONNET, A., HATON, J. P. & TRUONG-NGOC, J. M. (1986) *Systèmes-experts : Vers la maîtrise technique*. Informatique et Intelligence Artificielle, InterÉditions, Paris.
- BORILLO, A. (1996) « Exploration automatisée de textes de spécialité : repérage et identification automatique de la relation lexicale d'hyperonymie ». *Linx*, 34-35, pp.113-121.
- BOUAUD, J., HABERT, B., NAZARENKO, A. & ZWEIGENBAUM, P. (1997) « Regroupements issus de dépendances syntaxiques en corpus : catégorisation et confrontation à deux modélisations conceptuelles ». In *Actes des 1^{ères} journées "Ingénierie des Connaissances" (IC 1997)*, Roscoff, France, 20-22 mai 1997, pp.207-233.
- BOULANGER, J.-C. (1989) « L'évolution du concept de néologie de la linguistique aux industries de la langue ». In *Actes du colloque "Terminologie diachronique"*, C. De Schaetzen, Bruxelles, Belgique, 25-26 mars 1988, pp.193-211.
- BOULANGER, J. C. (1995) « Présentation : images et parcours de la socioterminologie ». *Meta, Presses de l'Université de Montréal*, XL(2), pp.194-205.

- BOURIGAULT, D. (2007) *Un analyseur syntaxique opérationnel : SYNTEX*. Habilitation à Diriger les Recherches, Université Toulouse 2.
- BOURIGAULT, D. & AUSSENAC-GILLES, N. (2003) « Construction d'ontologies à partir de textes ». In *Actes de la 10^{ème} conférence "Traitement Automatique des Langues Naturelles" (TALN 2003)*, Batz-sur-Mer, France, 11-14 juin 2003, pp.27-50.
- BOURIGAULT, D., AUSSENAC-GILLES, N. & CHARLET, J. (2004) « Construction de ressources terminologiques ou ontologiques à partir de textes : un cadre unificateur pour trois études de cas ». In J. M. Pierrel & M. Slodzian (Éd.), *Techniques Informatiques et structuration de terminologiques*, *Revue d'Intelligence Artificielle (RIA)*, Hermès, 18, Paris, pp.87-110.
- BOURIGAULT, D. & FABRE, C. (2000) « Approche linguistique pour l'analyse syntaxique de corpus ». *Cahiers de Grammaire, Université Toulouse 2*, 25, pp.131-151.
- BOURIGAULT, D., FABRE, C., FRÉROT, C., JACQUES, M.-P. & OZDOWSKA, S. (2005) « Syntex, analyseur syntaxique de corpus ». In *Actes de la 12^{ème} conférence "Traitement Automatique des Langues Naturelles" (TALN 2005)*, Dourdan, France, 6-10 juin 2005, pp.17-25.
- BOURIGAULT, D. & GALY, É. (2005) « Analyse distributionnelle de corpus de langue générale et synonymie ». In *Actes des 4^{èmes} Journées de la Linguistique de Corpus*, Lorient, France, 15-17 septembre 2005.
- BOURIGAULT, D. & JACQUEMIN, C. (2000) « Construction de ressources terminologiques ». In J. M. Pierrel (Éd.), *Industrie des langues*, Hermès, Paris, pp.215-233.
- BOURIGAULT, D. & LAMÉ, G. (2002) « Analyse distributionnelle et structuration de terminologie ». *Revue Traitement Automatique des Langues (TAL)*, 41(1), pp.1-10.
- BOURIGAULT, D. & SLODZIAN, M. (1999) « Pour une terminologie textuelle ». *Terminologies Nouvelles*, 19, pp.29-32.
- BOUTAYEB, S. (1996) « Constitution d'une terminologie multilingue : approche et méthode ». In *Actes des Journées Realiter "Réflexions méthodologiques sur le travail en terminologie et en terminotique dans les langues latines"*, H. Zinglé, Université de Nice-Sophia Antipolis, Nice, France.
- BOUTET, J. (1994) *Construire le sens*. Sciences pour la communication, Peter Lang, Bern.

- BOUTET, J. (2008) *La vie verbale au travail. Des manufactures aux centres d'appels*. Travail et activité humaine, Octares, Toulouse.
- BOUVERET, M. (1998) « Approche de la dénomination en langue spécialisée ». *Meta, Presses de l'Université de Montréal*, XLIII(3), pp.1-18.
- BOWKER, L. & PEARSON, J. (2002) *Working with Specialized Language: a Practical Guide to Using Corpora*. Routledge, London/New York.
- BRANCA-ROSOFF, S. (1999) « Des innovations et des fonctionnements de langue rapportés à des genres ». *Langage et société*, 87, pp.115-129.
- BRÉAL, M. (1897) *Essai de sémantique*. Gérard Monfort, Paris.
- BRUNOT, F. (1967) *Histoire de la langue française des origines à nos jours*. Armand Collin, Paris.
- BUITELAAR, P., CIMIANO, P. & MAGNINI, B. (Éd.) (2005) *Ontology Learning from Text: Methods, Evaluation and Applications*. IOS Press, Frontiers in Artificial Intelligence, 123.
- CABRÉ, M. T. (1998) *La terminologie : théories, méthodes et applications*. Armand Colin, Presses de l'Université d'Ottawa, Ottawa.
- CABRÉ, M. T. (2000) « Sur la représentation mentale des concepts ». In H. Béjoint & P. Thoiron (Éd.), *Le sens en terminologie*, Presses Universitaires de Lyon, Travaux du CRTT (Centre de Recherche en Terminologie et Traduction), Lyon, pp.20-39.
- CABRÉ, M.T. (2003) « Theories of Terminology: Their Description, Prescription and Explanation ». *Terminology*, 9(2), pp.163-199.
- CABRÉ, M. T. (2006) « NEOROM, réseau d'observatoires de la néologie des langues romanes ». *Neologica*, 1, pp.115-118.
- CABRÉ, M. T., DOMÈNECH, M., ESTOPÀ, R., FREIXA, J. & SOLÉ, É. (2003) « L'observatoire de néologie : conception, méthodologie, résultats et nouveaux travaux ». In J.F. Sablayrolles (Éd.), *L'innovation lexicale*, Honoré Champion, Paris, pp.125-147.
- CABRÉ, M. T. & YZAGURRI (DE), L. (1995) « Stratégie pour la détection semi-automatique des néologismes de presse ». *Traduction, Terminologie, Rédaction (TTR)*, 8(2), pp.89-100.

- CALBERG-CHALLOT, M., CANDEL, D. & FLEURY, S. (2006) « "Nucléaire" et "atomique", deux formes concurrentielles dans le domaine du nucléaire? ». In *Actes des 8^{èmes} Journées internationales d'Analyse statistique des Données Textuelles (JADT 2006)*, Besançon, France, 19-21 avril 2006, pp.223-234.
- CALBERG-CHALLOT, M., CANDEL, D. & ROCHE, C. (2007) « De la variation des usages au consensus terminologique : vers un dictionnaire de l'ingénierie nucléaire ». In *Actes de la conférence "Terminologie et Ontologie : Théories et applications" (Toth 2007)*, Annecy, France, 1er juin 2007, pp.119-141.
- CANDEL, D. (2004) « Wüster par lui-même ». In C. Cortès (Éd.), *Des fondements théoriques de la terminologie*, Cahiers du C.I.E.L. (Centre Interlangue d'Études en Lexicologie), Université Paris 7, pp.15-31.
- CARREÑO, I., LE SERREC, A. & BOUDREAU, M. (à paraître) « Evaluating Automatic Terminology Extraction Methods for the Construction of an Ontology ». In *Actes de la XV^{ème} conférence européenne "Language for Special Purposes" (LSP 2007)* (Éd.), Peter Lang, Hambourg, Allemagne, 27-31 août 2007.
- CARTIER, É. & SABLAYROLLES, J.-F. (2008) « Néologismes, dictionnaires et informatique ». *Cahiers de lexicologie*, 93(2), pp.175-192.
- CELOTTI, N. & MUSACCHIO, M. T. (2004) « Un regard diachronique en didactique des langues de spécialité ». *Revue de didactologie des langues-cultures et de lexiculturologie*, 135(3), pp.263-270.
- CHATEAU, C. (2007) « Drift and Shift: How "Continental" and "Continents" Move in Geological English ». In *Actes de la 28^{ème} conférence "International Computer Archive of Modern and Medieval English" (ICAME 2007)*, Stratford-upon-Avon, Royaume-Uni, 23-27 mai 2007.
- CHURCH, K., GALE, W., HANKS, P. & HINDLE, D. (1991) « Parsing, Word Associations and Typical Predicate-Argument Relations ». In M. Tomita (Éd.), *Current issues in parsing technology*, Kluwer Academic Publishers, Dordrecht, Pays-Bas, pp.103-112.
- CIMINO, J. J. (1996) « Formal Descriptions and Adaptative Mechanisms for Changes in Controlled Medical Vocabularies ». *Methods of Information in Medicine*, 35, pp.202-210.
- CIMINO, J. J., CLAYTON, P. D., HRIPCSAK, G. & JOHNSON, S. B. (1994) « Knowledge-based Approaches to the Maintenance of a Large Controlled Medical Terminology ». *Journal of the American Medical Informatics Association*, 1(1), pp.35-50.

- CNES-PRESSE (Dernière mise à jour : 2003) « SPOT 5 ». Consulté le 18 mai 2008 à l'adresse : http://www.cnes.fr/automne_modules_files/standard/public/p1778_bfdb95717dd168df28bd7ff8853b35f6spot5.pdf.
- CNES (2002) *Cours de techniques et technologies des véhicules spatiaux*. Cnes – Cépaduès Éditions.
- CNES (Dernière mise à jour : mars 2007) « Pharaon – Mesurer le temps avec une précision inégalée ». Consulté le 18 mai 2008 à l'adresse : <http://www.cnes.fr/web/4444-pharaon.php>
- CNES/CILF (1985) *Dictionnaire de Spatiologie*. Conseil international de la langue française (CILF), Paris.
- CNES/CILF (1992) *Dictionnaire de Spatiologie*. Conseil international de la langue française (CILF), Paris.
- CNES/CILF (1997) *Dictionnaire de Spatiologie*. Conseil international de la langue française (CILF), Paris.
- CNES/CILF (2001) *Dictionnaire de Spatiologie*. Conseil international de la langue française (CILF), Paris.
- COLLET, T. (1997) « La réduction des unités terminologiques complexes de type syntagmatique ». *Meta, Presses de l'Université de Montréal*, XLII(1), pp.193-206.
- COLLET, T. (2002) *La réduction des unités terminologiques complexes de type syntagmatique*. Thèse de doctorat en linguistique, Université de Montréal, Canada.
- COLLIER, A. (1993) « Issue of Large-Scale Collocational Analysis ». In B. Aarts, P. Haan (de) & N. Oostdijk (Éd.), *English Language Corpora: Design, Analysis and Exploitation*, Rodopi, Amsterdam, pp.289-298.
- CONDAMINES, A. (1994) « Terminologie et représentation des connaissances ». *La banque des mots*, 6 (Numéro spécial), pp.29-44.
- CONDAMINES, A. (1999) « Approche sémasiologique pour la constitution de Bases de Connaissances Terminologiques ». *Dyalang, Dynamiques sociolinguistiques*, Université de Rouen, 275, pp.101-117.

- CONDAMINES, A. (2000) « Les bases théoriques du groupe toulousain "Sémantique et Corpus" : ancrages et perspectives ». *Cahiers de Grammaire, "Sémantique et Corpus"*, 25, pp.5-28.
- CONDAMINES, A. (2003a) « Sémantique et corpus spécialisés : constitution de bases de connaissances terminologiques ». *Carnet de Grammaire, Rapports Internes de l'ERSS (Équipe de Recherche en Syntaxe et Sémantique), Habilitation à Diriger les Recherches, Toulouse 2*, 13.
- CONDAMINES, A. (2003b) « Vers la définition de genres interprétatifs ». In *Actes de la conférence "Terminologie et Intelligence Artificielle" (TIA 2003)*, Strasbourg, France, 31 mars – 1er avril 2003, pp.69-79.
- CONDAMINES, A. (2005) « Analyse linguistique de documents d'entreprises : demande appliquée et théorisation des phénomènes ». In D. Alamargot, P. Terrier & J.-M. Cellier (Éd.), *Production, compréhension et usages des écrits techniques au travail*, Octares, Paris pp.17-30.
- CONDAMINES, A. (2006) « Modes de construction du sens en corpus spécialisés ». *Cahiers de Grammaire "spécial anniversaire"*, 30, pp.75-88.
- CONDAMINES, A. (2007) « L'interprétation en sémantique de corpus : le cas de la construction de terminologies ». *Revue Française de Linguistique Appliquée*, XII(1), pp.39-52.
- CONDAMINES, A. & REBEYROLLE, J. (1997a) « Point de vue en langue spécialisée ». *Meta, Presses de l'Université de Montréal*, XLII(1), pp.174-184.
- CONDAMINES, A. & REBEYROLLE, J. (1997b) « Utilisation d'outils dans la constitution de bases de connaissances terminologiques : Expérimentation, limites, définition d'une méthodologie ». In *Actes des 1^{ères} Journées Scientifiques et Techniques FRANCIL de l'AUPELF-UREF*, Avignon, France, 15-16 avril 1997, pp.529-535.
- CONDAMINES, A. & REBEYROLLE, J. (2000) « Construction d'une base de connaissances terminologiques à partir de textes : expérimentation et définition d'une méthode ». In J. Charlet, G. Zacklad, G. Kassel & D. Bourigault (Éd.), *Ingénierie des connaissances, évolutions récentes et nouveaux défis*, Eyrolles, Paris, pp.225-241.
- CONDAMINES, A., REBEYROLLE, J. & SOUBEILLE, A. (2004) « Variation de la terminologie dans le temps : une méthode linguistique pour mesurer l'évolution de la connaissance en corpus ». In *Actes de la conférence "European Association for Lexicography" (EURALEX 2004)*, Lorient, France, 6 -10 juillet 2004, pp.547-557.

- CONEIN, B. (1985) « Une méthode descriptive en histoire est-elle possible? ». In P. Achard, M.P. Gruenais & D. Jauhin (Éd.), *Histoire et linguistique*, Éditions de la Maison des Sciences de L'Homme, Paris, pp.43-53.
- CORMIER, M. & RIOUX, L. P. (1991) « Procédés de formation et matrices terminogéniques en terminologie des systèmes experts ». *Meta, Presses de l'Université de Montréal*, XXXVI(1), pp.248-268.
- COSERIU, E. (1964) « Pour une sémantique diachronique structurale ». *Travaux de Linguistique et de Littérature*, II(1), pp.139-186.
- CROFT, W. (2000) *Explaining Language Change – An Evolutionary Approach*. Langman Linguistics Library.
- CUSIN-BERCHE, F. & MORTUREUX, M. F. (2004) « Autonymie et néologie ». In J. Authier-Revuz, M. Doury & S. Renoul-Touré (Éd.), *Parler des mots : le fait autonymique en discours*, Presses Sorbonne-Nouvelle, Paris, pp.113-122.
- DAILLE, B. (2003) « Conceptual Structuring through Term Variations ». In F. Bond, A. Korhonen, D. MacCarthy & A. Villacencio (Éd.), *Actes de la conférence "Association for Computational Linguistics" (ACL 2003), Atelier "Multiword Expressions : Analysis, Acquisition and Treatment"*, pp.9-16.
- DAILLE, B. (2005) « Variations and Application-Oriented Terminology Engineering ». *Terminology*, 11(1), pp.181-197.
- DAILLE, B., HABERT, B., JACQUEMIN, C. & ROYAUTÉ, J. (1996) « Empirical Observation of Term Variations and Principles for their Description ». *Terminology*, 3(2), pp.197-258.
- DARMESTER, A. (1886) *La vie des mots étudiée dans leurs significations*. Delagrave, Paris.
- DE SURMONT, J.-N. (2005) « Quelques réflexions métaterminologiques et propositions néologiques en terminologie ». *Le Langage et L'Homme*, XXXX(1), pp.123-140.
- DEJEAN, H. & GAUSSIER, E. (2002) « Une nouvelle approche à l'extraction de lexiques bilingues à partir de corpus comparables ». *Lexicometrica*, numéro spécial, J. Véronis (Éd.), Alignement lexical dans les corpus multilingues, pp.22-30.
- DELAPORTE, Y. (2005) « Origine et évolution des termes pour désigner les parents en langue des signes française ». *La langue des signes hier et aujourd'hui*.

- DELAVIGNE, V. (2001a) *Les mots du nucléaire. Contribution socioterminologique à une analyse des discours de vulgarisation*. Thèse de Sciences du Langage, Université de Rouen.
- DELAVIGNE, V. (2001b) « Repérage des termes dans un corpus de vulgarisation : aspects méthodologiques ». In *Actes de la conférence "Terminologie et Intelligence Artificielle" (TIA 2001)*, Nancy, France, 3-4 mai 2001, pp.33-43.
- DELESALLE, S. (1987) « Vie des mots et science des significations : Arsène Darmesteter et Michel Bréal ». *DRLAV, Revue de linguistique, "Dialogues du Marivaudage à la machine"*, 36-37, pp.265-314.
- DEPECKER, L. (2000) « Le signe, entre signifié et concept ». In H. Béjoint & P. Thoiron (Éd.), *Le sens en terminologie*, Presses Universitaires de Lyon, Travaux du CRTT (Centre de Recherche en Terminologie et Traduction, Lyon, pp.86-126.
- DEPECKER, L. (2002) *Entre signe et concept. Éléments de terminologie générale*, Presses Sorbonne Nouvelle, Paris.
- DEPECKER, L., MAMAVI, G., CHANSOU, M., GOUADEC, D., GUESPIN, L., ROUGES-MARTINEZ, J. & THOIRON, P. (Éd.) (1997) *La mesure des mots – Cinq études d'implantation terminologique*. Publications de l'Université de Rouen, Rouen.
- DESMET, I. (2002) « Néologie du portugais contemporain : une zone d'instabilité linguistique ». *Travaux et Documents, Presses Universitaires de Vincennes-Saint-Denis*, 16, pp.77-99.
- DESMET, I. (2007) « Éléments pour une théorie variationniste de la terminologie et des langues de spécialité ». *Cahiers du RIFAL, numéro spécial : terminologie, culture et société*, pp.3-13.
- DEVEREAUX FERGUSON, S. (2000) *Researching the Public Opinion Environment: Theories and Methods*. Thousand Oaks, Sage Publications, Londres, New Dehli.
- DROUIN, P. & DURY, P. (à paraître) « When Terms Disappear from a Specialized Lexicon: A Semi-Automatic Investigation into Necrology ». In *Actes de la conférence internationale "Language for Special Purposes" (LSP 2009)*, Aarhus, Danemark, 21-16 août 2009.
- DROUIN, P., MÉNARD, N. & PAQUIN, A. (2006) « Extraction semi-automatique des néologismes dans la terminologie du terrorisme ». In *Actes des 8^{èmes} Journées internationales d'Analyse statistique des Données Textuelles (JADT 2006)*, Besançon, France, 19-21 avril 2006, pp.389-400.

- DURY, P. (1997) *Étude comparative et diachronique de l'évolution*. Thèse de Linguistique – Lexicologie et Terminologie multilingues – Traduction, Université Louis-Lumière, Lyon 2.
- DURY, P. (1999) « Étude comparative et diachronique des concepts *ecosystem* et *écosystème* ». *Meta, Presses de l'Université de Montréal*, XLIV(3), pp.485-499.
- DURY, P. (2000) « Les variations sémantiques en terminologie : étude diachronique et comparative appliquée à l'écologie ». *Dyalang, Dynamiques sociolinguistiques, Presses Universitaires de Rouen*, 275, pp.17-32.
- DURY, P. (2002) « La terminologie diachronique, une aide à la traduction spécialisée, l'exemple de l'écologie ». In *Actes du II^{ème} Congrès International de Traduction Spécialisée (CITS)*, B. Promociones y Publicaciones Universitarias, Barcelone, Espagne, pp.201-209.
- DURY, P. (2004) « Building a Bilingual Diachronic Corpus of Ecology: The Long Road to Completion ». *ICAME Journal*, 28, pp.5-16.
- DURY, P. (2006) « Terminologie et mouvements diachroniques : un autre point de vue sur la variation ». In *Actes de la conférence "La terminologie scientifique et technique arabe à l'âge d'or"*, Lyon, France, 23-34 novembre 2006.
- DURY, P. (2007) « La déterminologisation du formant *eco-* et la terminologisation de l'adjectif *carbon neutral* : un aperçu diachronique de la migration des termes ». In *Actes de la conférence "Terminologie et Intelligence Artificielle" (TIA 2007)*, Sophia-Antipolis, France, 8-10 octobre 2007, pp.61-70.
- DURY, P. (2008a) « Les noms du pétrole : une approche diachronique de la métonymie onomastique ». *Lexis*, 1, « Polysemy / La polysémie », pp.10-22.
- DURY, P. (2008b) « The Rise of *Carbon Neutral* and *Compensation Carbone*: A Diachronic Investigation into the Migration of Vocabulary from the Language of Ecology to Newspaper Language and Vice Versa ». *Terminology*, 14(2), pp.230-248.
- DURY, P. & LERVAD, S. (2007) « La variation dans la terminologie de l'énergie : approches synchronique et diachronique, deux études de cas ». In *Actes du colloque "Terminologie : approches transdisciplinaires" (communication orale)*, Gatineau, Québec, Canada.

- DURY, P. & PICTON, A. (à paraître) « Terminologie et diachronie : vers une réconciliation théorique et méthodologique? ». *Revue Française de Linguistique Appliquée*.
- ECKARDT, R. (2003) « Meaning Change in Conceptual Montague Semantics ». In R. Eckardt, R. Von-Heusinger & C. Schwarze (Éd.), *Words in time: Diachronic Semantics from Different Points of View*, Mouton de Gruyter, Trends in Linguistics, Studies and Monographs, 143, Berlin/New York.
- ECKARDT, R., VON-HEUSINGER, R., AUTEUR & SCHWARZE, C. (Éd.) (2003) *Words in Time: Diachronic Semantics from Different Points of View*. Mouton de Gruyter, Trends in Linguistics, Studies and Monographs, 143, Berlin/New York.
- ENGWALL, G. (1994) « Not Chance but Choice: Criteria in Corpus Creation ». In S. Atkins & A. Zampolli (Éd.), *Computational Approaches to the Lexicon*, Oxford University Press, Oxford pp.49-82.
- ESTOPÀ BAJOT, R. (2001) « Les unités de signification spécialisées : élargissant l'objet du travail en terminologie ». *Terminology*, 7(2), pp.217-237.
- FABRE, C. & BOURIGAULT, D. (2008) « Exploiter des corpus annotés syntaxiquement pour observer le continuum entre arguments et circonstants ». *Journal of French Language Studies*, 18(1), pp.87-102.
- FABRE, C., HABERT, B. & LABBÉ, D. (1997) « La polysémie dans la langue générale et les discours spécialisés ». *Sémiotiques*, 13, pp.15-31.
- FAIRON, C. (2000) « Le système GlossaNet ». In C. Fairon (Éd.), *Le système INTEX, applications et développements. Actes des journées INTEX (1996 et 1999)*, John Benjamins, Amsterdam/Philadelphia.
- FAIRON, C. & COURTOIS, B. (2000) « Les corpus dynamiques et GlossaNet. Extension de la couverture lexicale des dictionnaires électroniques anglais ». In *Actes des 5^{èmes} Journées Internationales d'Analyse Statistique des Données Textuelles (JADT 2000)*, Lausanne, Suisse, 9-11 mars 2000.
- FAULSTICH, E. (1998/1999) « Principes formels et fonctionnels de la variation en terminologie ». *Terminology*, 5(1), pp.93-103.
- FERRET, O., GRAU, B., HURAUULT-PLANTET, M., ILLOUZ, G. & JACQUEMIN, C. (2001) « Utilisation des entités nommées et des variantes terminologiques dans un système

de question-réponse ». In *Actes de la 8^{ème} conférence "Traitement Automatique des Langues Naturelles"* (TALN 2001), Tours, France, 2-5 juillet 2001, pp.153-162.

FERRET, S. (1996) *Le bateau de Thésée – Le problème de l'identité à travers le temps*. Paradoxe, Les Éditions de Minuit, Paris.

FEUILLARD, C. (2007) « Grammaticalisation et synchronie dynamique ». *La linguistique*, 43(1), pp.3-28.

FEUILLARD, C. & KRIER, F. (2000) « Pourquoi et comment étudier le changement linguistique en synchronie ». *La Linguistique*, 36(1-2), pp.245-253.

FIALA, P. (1985) « La langue dans l'Histoire, l'Histoire dans la langue ». In P. Achard, M.P. Gruenais & D. Jauhin (Éd.), *Histoire et linguistique*, Éditions de la Maison des Sciences de L'Homme, Paris, pp.73-95.

FIRTH, J. R. (1957) *A Synopsis of Linguistic Theory, 1930-1955*. Studies in Linguistic Analysis, Philological Society.

FREIXA, J. (2002) *Anàlisi de la variació denominativa en textos de different grau d'especialització de l'àrea de medi ambient*. Thèse de doctorat en linguistique, Université de Barcelone.

FREIXA, J. (2006) « Causes of Denominative Variation in Terminology – A Typology Proposal ». *Terminology*, 12(1), pp.51-77.

GALARRETA, D. (2001) « Mesurer l'évolution des connaissances d'un projet spatial ». In *Actes de la conférence "Terminologie et Intelligence Artificielle"* (TIA 2001), Nancy, France, 3-4 mai 2001, pp.87-98.

GARDIN, B., LEFÈVRE, G., MARCELLES, M. & MORTUREUX, M. F. (1974) « À propos du sentiment néologique ». *Langages*, 36, pp.45-52.

GARGOURI, Y. (2003) *Maintenance d'ontologies de domaine à partir d'analyses textuelles*. Thèse de doctorat en informatique cognitive (Soutenance de projet de recherche), Université du Québec à Montréal (UQÀM), Montréal.

GARGOURI, Y. (2009) *Contribution à la maintenance des ontologies à partir d'analyses textuelles : extraction de termes et de relations entre termes*. Thèse de doctorat en informatique cognitive, Université du Québec à Montréal (UQÀM), Montréal.

- GARGOURI, Y., LEFEBVRE, B. & MEUNIER, J. G. (2004) « ONTOLOGICO : vers un outil d'assistance au développement itératif des ontologies ». In *Actes des Journées d'Études sur Terminologie, Ontologie et Représentation des connaissances (TERMINO'2004)*, Lyon, France.
- GASIGLIA, N. (2004) « Faire coopérer deux concordanciers-analyseurs pour optimiser les extractions en corpus ». *Revue Française de Linguistique Appliquée*, IX(1), pp.45-62.
- GAUDIN, F. (1993) *Pour une socioterminologie – Des problèmes sémantiques aux pratiques institutionnelles*. Publications de l'Université de Rouen, Rouen.
- GAUDIN, F. (1995) « Champs, clôtures et domaines : Des langues de spécialité à la culture scientifique ». *Meta*, XL(2), pp.229-238.
- GAUDIN, F. (2003) *Socioterminologie – Une approche sociolinguistique de la terminologie*. Champs linguistiques, De Boeck – Duculot, Bruxelles.
- GEERAERTS, D. (1991) « Grammaire cognitive et sémantique lexicale ». *Communications*, 53, pp.17-50.
- GEERAERTS, D. (2001) « Vers une onomasiologie diachronique quantitative ». In J. François (Éd.), *Théories contemporaines du changement sémantique*, Peeters, Leuven.
- GENTILHOMME, Y. (1994) « L'éclatement du signifié dans les discours technoscientifiques ». *Cahiers de lexicologie*, 64(1994), pp.5-35.
- GILLAM, L., TARIQ, M. & AHMAD, K. (2005) « Terminology and the Construction of Ontology ». *Terminology*, 11(1), pp.55-81.
- GOUADEC, D. (1990) *Terminologie – Constitution des données*. AFNOR Gestion, Paris.
- GREFENSTETTE, G. (1994) *Explorations in Automatic Thesaurus Discovery*. Kluwer Academic Publishers, Dordrecht, Pays-Bas.
- GROULT, M., LOUIS, P. & ROGER, J. (Éd.) (1988) *Transfert de vocabulaire dans les sciences*. Table ronde du CNRS, Histoire du vocabulaire scientifique, Paris.
- GRZEGA, J. (2002) « Some Aspects of Modern Diachronic Onomasiology ». *Linguistics*, 40(5), pp.1021-1045.

- GUESPIN, L. (1995) « La circulation terminologique et les rapports entre science, technique et production ». *Meta, Presses de l'Université de Montréal*, LX(2), pp.206-215.
- GUILBERT, L. (1965) *La formation du vocabulaire de l'aviation*. Thèse de Doctorat ès Lettres, Faculté des Lettres et Sciences Humaines, Université de Paris, Larousse, Paris.
- GUILBERT, L. (1967) *Le vocabulaire de l'astronautique*. Publications de l'Université de Rouen, Rouen.
- GUILBERT, L. (1973a) « La spécificité du terme scientifique et technique ». *Langue Française*, 17, pp.5-17.
- GUILBERT, L. (1973b) « Théorie du néologisme ». *Cahier de l'Association Internationale des Études Françaises*, 25, pp.9-29.
- GUILBERT, L. (Éd.) (1974) *La néologie lexicale*. Didier – Larousse, Langages, Paris.
- GUILBERT, L. (1975) *La créativité lexicale*. Collection "Langue et Langage", Larousse Université, Paris.
- HABERT, B. (2004) « Outiller la linguistique : de l'emprunt de techniques aux rencontres de savoirs ». *Revue Française de Linguistique Appliquée*, IX(1), pp.5-24.
- HABERT, B. (2005) *Instruments et ressources électroniques pour le français*. Collection "L'essentiel français", Ophrys, Paris.
- HABERT, B., FABRE, C. & ISSAC, F. (1998) *De l'écrit au numérique : constituer, normaliser et exploiter les corpus électroniques*. Interéditions, Masson, Paris.
- HABERT, B. & FUCHS, C. (2004) « Le traitement automatique des langues : des modèles aux ressources ». *Le Français Moderne*, « Traitements automatiques et ressources numérisées pour la français », LXXII(1), pp.1-13.
- HABERT, B., ILLOUZ, G. & FOLCH, H. (2004) « Dégroupier les sens : pourquoi, comment? ». In *Actes des 7^{èmes} Journées internationales d'Analyse statistique des Données Textuelles (JADT 2004)*, Louvain La Neuve, Belgique, 10-12 mars 2004, pp.565-576.
- HABERT, B. & NAZARENKO, A. (1996) « La syntaxe comme marche-pied de l'acquisition des connaissances : bilan critique d'une expérience ». In *Actes des Journées sur l'acquisition*

des connaissances, Association Française d'Intelligence Artificielle (AFIA), Sète, France, pp.137-142.

HABERT, B., NAZARENKO, A. & SALEM, A. (1997) *Les linguistiques de corpus*. Armand Colin, Paris.

HABERT, B. & ZWEIGENBAUM, P. (2003) « Classer les mots : sémantique à gros grains et méthodologie harrissienne ». *Revue de Sémantique et Pragmatique*, 13, pp.25-45.

HAMELIN, L.-E. (1995) « Le québécisme *nordicité* : de la néologie à la lexicalisation ». *Traduction, Terminologie, Rédaction*, 8(2), pp.51-65.

HAMON, T. (2000) *Variation sémantique en corpus spécialisé : Acquisition de relations de synonymie à partir de ressources lexicales*. Thèse de doctorat en informatique, Paris-Nord.

HAMON, T. & NAZARENKO, A. (2001a) « Exploitation de l'expertise humaine dans un processus de constitution de terminologie ». In *Actes de la conférence "Traitement Automatique des Langues" (TALN 2001)*, Tours, France, 2-5 juillet 2001, pp.213-222.

HAMON, T. & NAZARENKO, A. (2001b) « La structuration de terminologie : une nécessaire coopération ». In *Actes de la conférence "Terminologie et Intelligence Artificielle" (TIA 2001)*, Nancy, France, 3-4 mai 2001, pp.1-10.

HARRIS, Z. S. (1954) « Distributional Structure ». *Word*, 10(2-3), pp.146-162.

HARRIS, Z. S. (1988) *Language and Information*. Columbia University Press, New York.

HARRIS, Z. S. (1991) *A Theory of Language and Information. A Mathematical Approach*. Oxford University Press, Oxford.

HARRIS, Z. S., GOTTFRIED, M., RYCKMAN, T., MATTICK, P., DALADIER, A., HARRIS, T. & HARRIS, S. (1989) « The Form of Information in Science, Analysis of Immunology Sublanguage ». *Boston Studies in the Philosophy of Science*, Dordrecht, Pays Bas, Kluwer Academic Publishers, 104, pp.1-590.

HATON, J.-P., BOUZID, N., CHARPILLET, F., HATON, M.-C., LÂASRI, B., LÂASRI, H., MARQUIS, P., MONDOT, T. & NAPOL, A. (1991) *Le Raisonnement en intelligence artificielle : modèles, techniques et architectures pour les systèmes à base de connaissances*. Informatique intelligence artificielle, InterEditions, Paris.

- HEARST, M. (1991) « Noun Homograph Disambiguation Using Local Context in Large Corpora ». In *Actes de la 7^{ème} conférence annuelle "University of Waterloo Centre for the New OED and Text Research"*, Oxford, Royaume-Uni, pp.1-19.
- HEARST, M. (1992) « Automatic Acquisition of Hyponyms from Large Texts Corpora ». In *Actes du 15^{ème} colloque international en linguistique informatique (Coling 1992)*, Nantes, France, 23-28 août 1992, pp.539-545.
- HEARST, M. (1997) « TextTiling: Segmenting Text into Multi-Paragraph Subtopic Passages ». *Computational Linguistics*, 23(1), pp.33-64.
- HEIDEN, S. & LAVRENTIEV, A. (2004) « Ressources électroniques pour l'étude des textes médiévaux : approches et outils ». *Revue Française de Linguistique Appliquée*, IX(1), pp.99-118.
- HEIDEN, S. & PRÉVOST, S. (2002) « Étiquetage d'un corpus hétérogène de français médiéval : enjeux et modalités ». In C. D. Pusch & W. Raible (Éd.), *Romanistische Korpuslingustik : Korpora und gesprochene Sprache, Romance Corpus Linguistics : Corpora and Spoken Language (ScriptOraIia)*, Gunter Narr Verlag, 126, Tübingen, pp.127-136.
- HELGORSKY, F. (1981) « Les méthodes en histoire de la langue française. Évolution et stagnation ». *Le Français Moderne*, 49(2), pp.119-144.
- HO-DAC, L. M. (2007) *La position initiale dans l'organisation du discours : une exploration en corpus*. Thèse de Sciences du Langage, Université Toulouse 2.
- HO-DAC, L. M. & PÉRY-WOODLEY, M.-P. (2008) « Méthodologie exploratoire outillée pour l'étude de l'organisation du discours ». In *Actes du Congrès Mondial de Linguistique Française (CMLF 2008)*, Paris, France, 9-13 juillet 2008.
- HOENIGSWALD, H. (1963) « Are There Universals of Linguistic Change? ». In J. Greenberg (Éd.), *Universals of Language*, MIT Press, Cambridge, pp.30-52.
- HOLMES, J. (1994) « Inferring Language Change from Computer Corpora: Some Methodological Problems ». *ICAME Journal*, 18, pp.27-40.
- HOUEBINE, A. M. (1985) « Pour une linguistique synchronique dynamique ». *La Linguistique*, 21, pp.7-36.

- HOWELL, D. C. (2008) *Statistiques en sciences humaines* (2^{nde} édition). Ouvertures psychologiques, De Boeck, Bruxelles.
- HUMBLEY, J. (1994) « Quelques aspects de la datation de termes techniques : le cas de l'enregistrement et de la reproduction sonores ». *Meta, Presses de l'Université de Montréal*, XXXIX(4), pp.701-715.
- HUMBLEY, J. (2003) « La néologie en terminologie ». In J.-F. Sablayrolles (Éd.), *L'innovation lexicale*, Honoré Champion, Paris, pp.261-278.
- HUMBLEY, J. (2004) « La réception de l'œuvre d'Eugen Wüster dans les pays de langue française ». *Cahiers du C.I.E.L. (Centre Interlangue d'Études en Lexicologie)*, Université Paris 7, pp.33-51.
- HUMBLEY, J. (2006) « La néologie : interface entre ancien et nouveau ». In R. Greenstein (Éd.), *Langues et cultures : une histoire d'interface*, Publications de la Sorbonne, Paris, pp.91-103.
- HUMBLEY, J. (2007) « La néonymie est-elle un concept opératoire en terminologie? ». In *Actes de la conférence "Terminologie : approches transdisciplinaires" (communication orale)*, Gatineau, Québec, Canada, 2-4 mai 2007.
- HUMBLEY, J. (à paraître) *La néonymie : la néologie terminologique*. Champion, Paris.
- IBEKWE-SANJUAN, F. (1998) « Terminological Variation, a Means of Identifying Research Topics from Texts ». In *Actes de la "Joint International Conference on Computational Linguistics" (ACL-COLING'98)*, Montréal, Canada, 10-14 août 1998, pp.564-570.
- IBEKWE-SANJUAN, F. (2005) « Repérage et annotation d'indices de nouveautés dans les écrits scientifiques ». In *Actes du colloque "Indice, Index, Indexation"*, Université Charles de Gaulle, Lille, France, 3-4 novembre 2005, pp.261-276.
- IBEKWE-SANJUAN, F. (2007) *Fouille de textes. Méthodes, outils et applications*. Hermès – Lavoisier, Paris.
- IBEKWE-SANJUAN, F. & SANJUAN, E. (2002) « From Term Variants to Research Topics ». *Knowledge Organisation*, 29(3/4), pp.181-197.
- IBEKWE-SANJUAN, F. & SANJUAN, E. (2004) « Mapping the Structure of Research Topics through Term Variant Clustering: The TermWatch System ». In *Actes des 7^{èmes} Journées*

internationales d'Analyse statistique des Données Textuelles (JADT 2004), Louvain La Neuve, Belgique, 10-12 mars 2004, pp.589-600.

IEZZI, D. F., AURELI, E. & CRISTOFORI, F. (2008) « On-Line Forums: A Window on Language Evolution ». In *Actes des 9^{èmes} Journées internationales d'Analyse statistique des Données Textuelles (JADT 2008)*, Lyon, France, pp.561-568.

JACQUART, H. (1988) *Qui? Quoi? Comment? Ou La Pratique des sondages*. Eyrolles, Paris.

JACQUEMIN, C. (1997) *Variation terminologique : reconnaissance et acquisition automatique de termes et de leurs variantes en corpus*. Habilitation à Diriger des Recherches, Université de Nantes.

JACQUEMIN, C. (2001) *Spotting and Discovering Terms through NLP*. MIT Press, Cambridge MA.

JACQUES, M.-P. (2003) *Approche en discours de la réduction des termes complexes dans les textes spécialisés*. Thèse de doctorat en Sciences du Langage, Université Toulouse 2, Toulouse.

JACQUES, M.-P. (2005) « L'emploi de termes réduits comme révélateur de la centralité dans le domaine ». In *Actes des 7^{èmes} journées scientifiques du Réseau "Lexicologie, Terminologie, Traduction" (Agence Universitaire de la Francophonie) : « Mots, termes et contextes »*, Bruxelles, Belgique, 8-10 septembre 2005, pp.299-308.

JANSSEN, M. (2008) « NeoTrack – Un analyseur de néologismes en ligne ». In *Actes du 1^{er} Congrès International de Néologie des langues romanes (Cinéo 2008)*, Barcelone, Espagne, 07-10 mai 2008.

JOSSELIN-LERAY, A. (2005) *Place et rôle des terminologies dans les dictionnaires généraux unilingues et bilingues. Étude d'un domaine de spécialité : volcanologie*. Thèse de doctorat en Linguistique – Lexicologie et Terminologie multilingues – Traduction, Université Louis-Lumière, Lyon 2.

KAGEURA, K. (1999) « On the Study of Dynamics of Terminology: A Proposal of a Theoretical Framework ». *Research Bulletin of the National Center for Science Information System*, 11, pp.19-26.

KAGEURA, K. (2002) *The Dynamics of Terminology: A Descriptive Theory of Term Formation and Terminological Growth*. Terminology and Lexicography Research and Practice, John Benjamins Publishing Company, Amsterdam, Pays-Bas.

- Kilgarriff, A. (2006) « Collocationality (and How to Measure it) ». In *Actes de la 12^{ème} conférence internationale "European Association for Lexicography" (EURALEX 2006)* Turin, Italie, pp.6-9 septembre 2006.
- KLEIN, M. (2004) *Change Management for Distributed Ontologies*. Thèse de doctorat de l'Université Libre d'Amsterdam, Pays-Bas.
- KOCH, P. (2001) « Pour une approche cognitive du changement sémantique lexical : aspect onomasiologique ». In J. François (Éd.), *Théories contemporaines du changement sémantique*, Peeters, Leuven, pp.75-96.
- KOCOUREK, R. (1982) *La langue française de la technique et de la science*. Brandstetter Verlag, Zurich.
- KOCOUREK, R. (1991a) *La Langue Française de la Technique et de la Science* (2^{ème} édition). Brandstetter Verlag, Zurich.
- KOCOUREK, R. (1991b) « Textes et termes ». *Meta*, 36(1), pp.71-76.
- KRISTEVA, J. (1969) *Sèmiôtikè. Recherches sur une sémanalyse*. Seuil, Paris.
- KUHN, A. (1999) *La structure des révolutions scientifiques*. Champs, Flammarion, Paris.
- KYTÖ, M., RUDANKO, J. & SMITTERBERG, E. (2000) « Building a Bridge Between the Present and the Past: A Corpus of 19th-Century English ». *ICAME Journal*, 24, pp.85-97.
- L'HOMME, M.-C. (1998) « Le statut du verbe en langue de spécialité et sa description lexicographique ». *Cahiers de lexicologie*, 73(2), pp.61-84.
- L'HOMME, M.-C. (2000) « Les enseignements d'un mot polysémique sur les modèles de la terminologie ». *Cahiers de Grammaire*, 25, pp.71-91.
- L'HOMME, M.-C. (2001) « Nouvelles technologies et recherche terminologique. Techniques d'extraction des données terminologiques et leur impact sur le travail du terminographe ». In *Actes de la conférence "L'impact des nouvelles technologies sur la gestion terminologique"*, Toronto, Université de York.
- L'HOMME, M.-C. (2002) « What can Verbs and Adjectives Tell us about Terms? ». In *Actes de la conférence internationale "Terminology and Knowledge Engineering" (TKE 2002)*, Nancy, France, 28-30 août 2002, pp.65-70.

- L'HOMME, M.-C. (2004a) « A Lexico-semantic Approach to the Structuring of Terminology ». In *Actes de la conférence Computerm 2004 (Coling 2004)*, Genève, Suisse, 29 août 2004, pp.7-14.
- L'HOMME, M.-C. (2004b) *Terminologie : principes et techniques*. Presses de l'Université de Montréal, Montréal.
- L'HOMME, M.-C. (2005) « Sur la notion de "terme" ». *Meta, Presses de l'Université de Montréal*, L(4), pp.1112-1132.
- L'HOMME, M.-C. (2008) « Ressources lexicales, terminologiques et ontologiques : une analyse comparative dans le domaine de l'informatique ». *Revue française de linguistique appliquée*, 13(1), pp.97-118.
- L'HOMME, M.-C., HEID, U. & SAGER, J. C. (2003) « Terminology during the past decade (1994-2004) : An Editorial statement ». *Terminology*, 9(2), pp.151-161.
- L'HOMME, M.-C. & MARSHMAN, E. (2006) « Terminological Relationships and Corpus-based Methods for Discovering them: An Assessment for Terminographers" ». In L. Bowker (Éd.), *Lexicography, Terminology, and Translation. Text-based studies in honour of Ingrid Meyer*, University of Ottawa Press, Ottawa, pp.67-80.
- LAIGNELET, M. (2007) « La recherche d'information évolutive dans des documents de type encyclopédique : l'apport de techniques linguistiques ». In *Actes de la Conférence en Recherche d'Information et Applications (CORIA 2007)*, Saint-Étienne, France, 28-30 mars 2007, pp.449-454.
- LAIGNELET, M. (2009) *Associer analyse syntaxique et analyse discursive pour le repérage d'informations obsolètes dans les documents encyclopédiques*. Thèse de doctorat en Sciences du Langage, Toulouse 2, Toulouse.
- LAIGNELET, M. & PIMM, C. (2007) « Utiliser une segmentation thématique TextTiling pour le repérage de segments d'information évolutive dans un corpus de textes encyclopédiques ». In *Actes de la 14^{ème} conférence "Traitement Automatique des Langues Naturelles" (TALN'2007) (communication affichée)*, Toulouse, France, 5-8 juin 2007.
- LASS, R. (1980) *On Explaining Language Change*. Cambridge University Press, Cambridge.
- LASS, R. (2000) *Historical Linguistics and Language Change*. Cambridge Studies in Linguistics, Cambridge.

- LEBEL, M. É. (2003) *Le morphème par en Ancien Français*. Thèse de Linguistique, Université du Québec à Montréal (UQÀM), Montréal.
- LEBLANC, J.-M. & MARTINEZ, W. (2006) « L'analyse contrastive des réseaux de cooccurrence. Le monde dans les discours des présidents de la Cinquième République ». In *Actes des 8^{èmes} Journées internationales d'Analyse statistique des Données Textuelles (JADT 2006)*, Besançon, France, 19-21 avril 2006.
- LELUBRE, X. (1992) *La terminologie arabe contemporaine de l'optique : faits, théorie, évolution*. Thèse, Université Louis-Lumière, Lyon 2.
- LERAT, P. (1989) « Les fondements théoriques de la terminologie ». *La banque des mots*, Numéro spécial, pp.51-62.
- LERAT, P. (1995) *Les langues spécialisées*. Presses Universitaires de France, Paris.
- LEROY, S. (2004) « Extraire sur patrons : allers et retours entre analyse linguistique et repérage automatique ». *Revue Française de Linguistique Appliquée*, IX(1), pp.25-43.
- LINO, M. T. (2005) « Contextes et néologie terminologique dans le domaine médical ». In *Actes des 7^{èmes} journées scientifiques du réseau "Lexicologie, Terminologie, Traduction" (Agence Universitaire de la Francophonie) : « Mots, termes et contextes »*, Bruxelles, Belgique, 8-10 septembre 2005, pp.509-514.
- LOSEE, R. M. (1995) « The Development and Migration of Concepts from Donor to Borrower Disciplines: Sublanguage Term Use in Hard and Soft Sciences ». In *Actes de la 5^{ème} conférence internationale "Scientometrics and Informetrics"*, Chicago, États-Unis, 13 septembre 1995, pp.265-274.
- LOSEE, R. M. (1998) *Text Retrieval and Filtering: Analytic Models of Performance*. Information Retrieval., Kluwer Academic Publishers, Dordrecht, Pays Bas.
- LOZACHMEUR, G. (2007) « Diachronie des textes fondateurs de l'unification linguistique française du XVI^{ème} au XVIII^{ème} siècle ». In *Actes des 7^{èmes} Journées de l'ERLA (Équipe de Recherche en Linguistique Appliquée)*, Brest, France, 17-18 novembre 2006.
- LUONG, P. (2007) *Gestion de l'évolution d'un web sémantique d'entreprise*. Thèse de doctorat de l'École des Mines, Paris.

- MAEDCHE, A. & STAAB, S. (2001) « Ontology Learning for the Semantic Web ». *IEEE Intelligent Systems*, 16(2), pp.72-79.
- MAIR, C. (1997) « Parallel Corpora: A Real-Time Approach to the Study of Language Change in Progress ». In M. Ljung (Éd.), *Corpus-Based Studies in English*, GA-Rodopi, Amsterdam & Atlanta, pp.195-209.
- MAIR, C. (2002) « Three Changing Patterns of Verb Complementation in Late Modern English: A Real-time Study Based on Matching Text Corpora ». *English Language and Linguistics*, Cambridge University Press, 6(1), pp.105-131.
- MAIR, C. & HUNDT, M. (Éd.) (2000) *Corpus Linguistics and Linguistic Theory*. Rodopi, Language and Computers: Studies in Practical Linguistics, 33, Amsterdam/Atlanta.
- MAIR, C. & LEECH, G. (2006) « Current Change in English Syntax ». In B. Aarts & A. MacMahon (Éd.), *The Handbook of English Linguistics*, Blackwell, Oxford, pp.318-342.
- MALAISÉ, V., ZWEIGENBAUM, P. & BACHIMONT, B. (2005) « Mining Defining Contexts to Help Structuring Differential Ontologies ». *Terminology* 11(1), pp.21-53.
- MARCHELLO-NIZIA, C. (1995) *L'évolution du français – Ordre des mots, démonstratifs, accent tonique*. Linguistique, Armand Colin, Paris.
- MARCHELLO-NIZIA, C. (2002) « A NLP-driven Approach to Historical Linguistics ». In C. D. Pusch, J. Kabatek & W. Raible (Éd.), *Romanistische Korpuslinguistik II, Korpora und diachrone Sprachwissenschaft – Romance Corpus Linguistics II, Corpora and Diachronic Linguistics*, Gunter Narr Verlag, Tübingen, pp.11-30.
- MARCHELLO-NIZIA, C. (2004) « Linguistique historique, linguistique outillée : les fruits d'une tradition ». *Le Français Moderne*, 72(1), pp.58-70.
- MARCHELLO-NIZIA, C. (2006) *Grammaticalisation et changement linguistique*. Champs Linguistiques – Recherche, De Boeck, Bruxelles.
- MARSHMAN, E. (2006) *Lexical Knowledge Patterns for Semi-automatic Extraction of Cause-effect and Association Relations from Medical Texts: A Comparative Study of English and French*. Thèse de doctorat en Traduction, option Terminologie, Université de Montréal, Canada.

- MARTINET, A. (1990) « La synchronie dynamique ». *La Linguistique, Presses Universitaires de France*, 26(1), pp.13-23.
- MATHIEU, Y. (1998) « Étude pour une extraction automatique de néologismes ». In A. Clas, S. Mejri & T. Baccouche (Éd.), *Actes des 5^{èmes} journées scientifiques du Réseau "Lexicologie, Terminologie, Traduction"* (Agence Universitaire de la Francophonie) : « La mémoire des mots », Tunis, Tunisie, pp.455-459.
- MATHIEU, Y., GROSS, G. & FOUQUERÉ, C. (1998) « Vers une extraction automatique de néologismes ». *Cahiers de lexicologie*, 72(1), pp.199-208.
- MAYAFFRE, D. (2005) « Rôle et place des corpus en linguistique : Réflexions introductives ». In *Actes des Journées d'Etude TOUlousaines (JETOU 2005)*, « Rôle et place des corpus en linguistique », P. Vergely, Toulouse, France, 1-2 juillet 2005, pp.5-17.
- MCENERY, T. & WILSON, A. (2001) *Corpus Linguistics (seconde édition)*. Edinburgh University Press, Edinburgh, Edimbourg, Royaume-Uni.
- MEILLET, A. (1905) « Comment les mots changent de sens ». *L'Année sociologique*, 9, pp.1-38.
- MEJRI, S. (2005) « La reconnaissance automatique des néologismes de sens ». In *Actes des 7^{èmes} journées scientifiques du réseau "Lexicologie, Terminologie, Traduction"* (Agence Universitaire de la Francophonie) : « Mots, termes et contextes », Bruxelles, Belgique, 8-10 octobre 2005, pp.545-558.
- MEL'ČUK, I. A, CLAS, A & POLGUÈRE, A (1995) *Introduction à la lexicologie explicative et combinatoire*. Duculot, Paris.
- MÉLA, A. (2004) « Linguistes et 'talistes' peuvent coopérer : repérage et analyse des gloses ». *Revue Française de Linguistique Appliquée*, IX(1), pp.63-82.
- MEURMAN-SOLIN, A. (1995) « A New Tool : The Helsinki Corpus of Older Scots (1450-1700) ». *ICAME Journal*(19), pp.49-62.
- MEYER, I. (2001) « Extracting Knowledge-Rich Contexts for Terminography : A Conceptual and Methodological Framework ». In D. Bourigault, M.C. L'Homme & C. Jacquemin (Éd.), *Recent Advances in Computational Terminology*, John Benjamins Publishing Company, Amsterdam/New York, pp.279-302.

- MEYER, I. & MACKINTOSH, K. (2000) « "L'étirement" du sens terminologique : aperçu du phénomène de la déterminologisation ». In H. Béjoint & P. Thoiron (Éd.), *Le sens en terminologie*, Presses Universitaires de Lyon, Travaux du CRTT (Centre de Recherche en Terminologie et Traduction), Lyon, pp.198-217.
- MEYER, I., MACKINTOSH, K., BARRIÈRE, C. & MORGAN, T. (1999) « Conceptual Sampling for Terminographical Corpus Analysis ». In *Actes de la conférence internationale "Terminology and Knowledge Engineering" (TKE 1999)*, Innsbruck, Autriche, 23-27 août 1999, pp.256-267.
- MINETT, J. & WANG, W. S. Y. (2003) « On Detecting Borrowing: Distance-based and Character-based Approaches ». *Diachronica*, 20(2), pp.289-330.
- MØLLER, B. (1994) « Où va la terminologie technique française? ». *La Banque des mots*, 6 (Numéro spécial), pp.71-98.
- MØLLER, B. (1998) « À la recherche d'une terminochronie ». *Meta*, Presses de l'Université de Montréal, XLIII(3), pp.426-453.
- MORIN, E., DUFOUR-KOWALSKI, S. & DAILLE, B. (2004) « Extraction de terminologies bilingues à partir de corpus comparables ». In *Actes de la 11^{ème} conférence "Traitement Automatique des Langues Naturelles" (TALN 2004)*, Fès, Maroc, 19-21 avril 2004, pp.103-122.
- MULLER, C. (1968) *Initiation à la statistique lexicale*. Langue et Langage, Larousse, Paris.
- NICOLAS, A. (1980) « "Sélection naturelle" et synonymie ». *Langue Française*, 48, pp.89-99.
- NORMAND, C. (2000) *Saussure. Figures du savoir*, Les belles lettres, Paris.
- NYCKEES, V. (1998) *La Sémantique*. Collection Sujets, Belin, Paris.
- NYCKEES, V. (2001) « Changement de sens et déterminisme socio-culturel ». In J. François (Éd.), *Théories contemporaines du changement sémantique*, Peeters, Leuven.
- OLIVER, D. E., SHAHAR, Y., MUSEN, M. & SHORTLIFFE, E. H. (1999) « Representation of Change in Controlled Medical Terminologies ». *Artificial Intelligence in Medicine*, 15(1), pp.53-76.

- OLLINGER, S. & VALETTE, M. (2008) « La créativité lexicale : des pratiques sociales aux textes ». In *Actes du 1^{er} Congrès International de Néologie des langues romanes (Cinéo) (communication affichée)*, Barcelone, Espagne, 07-10 mai 2008.
- OTMAN, G. (1993) « La talon d'Achille de l'expertise terminologique ». *La Banque des mots*, numéro spécial 5, pp.75-86.
- OTMAN, G. (1996) *Les Représentations sémantiques en terminologie*. Masson, Paris.
- OTTENS, K., AUSSÉNAC-GILLES, N., GLEIZES, M. P. & CAMPS, V. (2007) « Dynamic Ontology Co-Evolution from Texts: Principles and Case Study ». In *Actes de l'Atelier International "Emergent Semantics and Ontology Evolution", 6^{ème} Conférence Internationale sur le Web Sémantique (ISWC)*, Busan, Corée, 12 novembre 2007, pp.70-83.
- PAQUIN, A. (2007) *Étude de la néologie dans la terminologie du terrorisme avant et après septembre 2001 : une approche lexicométrique*. Mémoire de maîtrise, Université de Montréal.
- PEARSON, J. (1998) *Terms in Context*. John Benjamins Publishing Company, Amsterdam/Philadelphia.
- PÉRY-WOODLEY, M.-P. (1995) « Quels corpus pour quels traitements automatiques? ». *Revue TAL (Traitement Automatique des Langues)*, 36(1-2), pp.213-232.
- PÉRY-WOODLEY, M.-P. (2000) « Une pragmatique à fleur de texte : approche en corpus de l'organisation textuelle ». *Rapports Internes de l'ERSS (Équipe de Recherche en Syntaxe et Sémantique), Habilitation à Diriger les Recherches, Toulouse 2*, 8.
- PÉRY-WOODLEY, M.-P. (2005) « Discours, corpus, traitements automatiques ». In A. Condamines (Éd.), *Sémantique et corpus*, Hermès, Paris, pp.177-210.
- PETIT, G. (2001) « L'introuvable identité du terme technique ». *Revue française de linguistique appliquée*, 1(2), pp.63-79.
- PICTON, A. (2007) « "Migration" des connaissances dans le temps : apport d'une analyse automatique des dépendances des termes ». In *Actes de la conférence "Terminologie et Intelligence Artificielle" (TIA 2007)*, Sophia-Antipolis, 8-9 octobre 2007.
- PICTON, A. (2008) « Combining Clues to Explore Knowledge Evolution in Texts ». In *Actes de la conférence internationale "Terminology and Knowledge Engineering" (TKE 2009)*, Copenhagen, Danemark, 18-21 août 2008.

- POIBEAU, T. (2005) « Parcours interprétatifs et terminologie ». In *Actes de la conférence "Terminologie et Intelligence Artificielle" (TIA 2005)*, Rouen, France.
- PORTELANCE, C. (1987) « Fertilisation terminologique ou insémination terminologique artificielle ? ». *Meta, Presses de l'Université de Montréal*, XXXII(3), pp.356-360.
- PORTELANCE, C. (1996) « De la nomination : catégorisation et syntagmatique ». *Sémiotique appliquée/Applied Semiotics, numéro spécial "Le référent"*, 1(2), pp.55-63.
- PRÉVOST, S. (2005) « Constitution et exploitation d'un corpus de français médiéval : enjeux, spécificités et apports ». In A. Condamines (Éd.), *Sémantique et Corpus*, Hermès, Paris, pp.147-176.
- PRÉVOST, S. (2008) « Corpus informatisés de français médiéval : contraintes sur leur constitution et spécificités de leurs apports ». *Corpus, "Constitution et exploitation des corpus d'ancien et de moyen français" [En ligne]*, 7, mis en ligne le 23 juin 2009, consulté le 10 juillet 2009 à l'adresse <http://corpus.revues.org/index1500.html>.
- PROST, A. (1996) *Douze leçons sur l'histoire*. Points Histoire, Éditions du Seuil, Paris.
- PRUVOST, J. & SABLAYROLLES, J.-F. (2003) *Les néologismes. Que sais-je?*, Presses Universitaires de France (PUF), Paris.
- QUEMADA, B. (1955) *Introduction à l'étude du vocabulaire médical, 1600-1710*. Les Belles Lettres, Paris.
- QUEMADA, B. (1978) « Technique et Langage, la formation des vocabulaires français des techniques ». In B. Gille (Éd.), *Histoire des techniques, techniques et civilisations, techniques et sciences*, Gallimard, Paris, pp.1146-1240.
- QUIRION, J. (2003a) *La mesure de l'implantation terminologique : proposition d'un protocole*. Collection Langue et Société, Office Québécois de la Langue Française, Montréal.
- QUIRION, J. (2003b) « Methodology for the Design of a Standard Research Protocol for Measuring Terminology Usage ». *Terminology*, 9(1), pp.29-49.
- QUIRION, J. (2004) « État de la question sur la nature des facteurs d'implantation terminologique ». In D. Gouadec (Éd.), *Mondialisation, localisation, francophonie(s)*, La Maison du Dictionnaire, Paris, pp.193-200.

- RASTIER, F. (1991) *Sémantique et recherches cognitives*. Presses Universitaires de France, PUF, Paris.
- RASTIER, F. (1995) « Le terme : entre ontologie et linguistique ». *La banque des mots*, 7, pp.35-65.
- RASTIER, F. (2000) « De la sémantique cognitive à la sémantique diachronique : les valeurs et l'évolution des classes lexicales ». In J. François (Éd.), *Théories contemporaines du changement linguistique*, Peeters, Leuven, pp.135-164.
- RASTIER, F. (2001) *Arts et sciences du texte*. Formes sémiotiques, Presses Universitaires de France, PUF, Paris.
- RASTIER, F. (2005) « Sémiotique du cognitivisme et sémantique cognitive : Questions d'histoire et d'épistémologie. mars 2005 [en ligne]. ». *Revue Texto ! [en ligne]*, consulté le 24 novembre 2006 à l'adresse <http://www.revue-texto.net>.
- RAVAOARIMALALA, É. (2005) « Moderniser sa langue pour mieux apprendre les langues étrangères ». In *Actes des 7^{èmes} journées scientifiques du réseau "Lexicologie, Terminologie, Traduction"* (Agence Universitaire de la Francophonie) : « Mots, termes et contextes », Bruxelles, Belgique, 8-10 septembre 2005, pp.535-544.
- REBEYROLLE, J. (1996) *Polysémie dans les langues spécialisées : une question de point de vue?* Mémoire de DEA, Université Toulouse 2.
- RENOUF, A. (1993) « A Word in Time: First Findings from the Investigation of Dynamic Texts ». In B. Aarts, P. Haan (de) & N. Oostdijk (Éd.), *English Language Corpora: Design, Analysis and Exploitation*, Rodopi, Amsterdam, pp.279-288.
- RENOUF, A. (1994) « Corpora and Historical Dictionaries ». In I. Lancashire & T. Russon Wooldridge (Éd.), *Early Dictionary Databases*, Center for Computing in the Humanities, University of Toronto, Toronto, pp.219-235.
- RENOUF, A. (2002) « The Time Dimension in Modern English Corpus Linguistics ». In *Actes de la 4^{ème} Conférence Internationale "Teaching and Learning by Doing Corpus Analysis"*, B. Kettemann & G. Marko (Éd.), Language and Computers, Graz, France, 19-24 juillet 2000, pp.27-41.
- RENOUF, A. & KEHOE, A. (Éd.) (2006) *The Changing Face of Corpus Linguistics*. Rodopi, Language and Computers (55), Amsterdam-New York.

- REY, A. (1976) « Néologisme : un pseudo-concept? ». *Cahiers de lexicologie*, 28(1), pp.3-17.
- REY, A. (1979) *La terminologie : noms et notions*. Que sais-je?, Presses Universitaires de France (PUF), Paris.
- REY, A. (1988) « Les fonctions de la terminologie : du social au théorique ». *Office de la langue française, Société des traducteurs du Québec (OLF-STQ)*, pp.87-108.
- REYMOND, D. (2002) « Méthodologie pour la création d'un dictionnaire distributionnel dans une perspective d'étiquetage lexical semi-automatique ». In *Actes de la 9^{ème} Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL 2002)*, Nancy, France, 24-27 juin 2002, pp.405-414.
- ROCHE, S. & BOWKER, L. (1999) « Cenit : Système de détection semi-automatique des néologismes ». *Terminologies Nouvelles*, 20, pp.12-16.
- ROGOZAN, D. (2008) *Gestion de l'évolution des ontologies : méthodes et outils pour un référencement sémantique évolutif fondé sur une analyse des changements entre versions d'ontologie*. Thèse de doctorat de l'université du Québec à Montréal, Montréal, Canada.
- RONDEAU, G. (1981) *Introduction à la terminologie*. Gaëtan Morin, Montréal.
- RONDEAU, G. & FELBER, H. (Éd.) (1981) *Textes choisis de terminologie*. GIRSTERM, Université de Laval, Québec.
- ROTHENBURGER, B. & GALARRETA, D. (2004) « Facing Knowledge Evolution in Space Projects: A New Way of Thinking About ». In *Actes du 55^{ème} congrès international de l'astronautique*, Vancouver, Canada, 4-8 octobre 2004, pp.1301-1310.
- RYCKMAN, T. (1990) « De la structure d'une langue aux structures de l'information dans le discours et dans les sous-langages scientifiques ». *Langages* (A. Daladier (dir.)), 99, pp.21-38.
- SABLAYROLLES, J. F. (1996a) « Néologisme et nouveauté(s) ». *Cahiers de lexicologie*, 69, pp.5-42.
- SABLAYROLLES, J. F. (1996b) « Néologismes : une typologie des typologies ». *Cahiers du CIEL*, 1996-1997, pp.11-48.

- SABLAYROLLES, J. F. (1997) « Locutions néologiques ». In P. Fiala, P. Lafon & M. F. Piguet (Éd.), *La locution : entre lexique, syntaxe et pragmatique. Identification en corpus, traitement, apprentissage*, Publications de l'INaLF, Collection Saint-Cloud, Klincksieck, Paris, pp.321-331.
- SABLAYROLLES, J. F. (2000a) *La néologie en français contemporain. Examen du concept et analyse de productions néologiques récentes*. Honoré Champion, Paris.
- SABLAYROLLES, J. F. (2000b) « Lexique et processus ». *Cahiers de Grammaire*, 77(2), pp.5-26.
- SABLAYROLLES, J. F. (2002a) « Des néologismes au signifiant très significatif ». In C. Gruaz (Éd.), *Quand le mot fait signe. Pour une sémiotique de l'écrit*, Dyalang, Dynamiques sociolangagières, Publications de l'Université de Rouen, Rouen, pp.97-117.
- SABLAYROLLES, J. F. (2002b) « Fondements théoriques des difficultés pratiques du traitement des néologismes ». *Revue française de linguistique appliquée*, VII(1), pp.97-111.
- SABLAYROLLES, J. F. (2003a) « Le sentiment néologique ». In J. F. Sablayrolles (Éd.), *L'innovation lexicale*, Honoré-Champion, Paris, pp.279-295.
- SABLAYROLLES, J.F. (Éd.) (2003b) *L'innovation lexicale*. Honoré Champion, Paris.
- SADER FEGHALI, L. (2005) « La presse vue à travers Néoscope : quand les contextes médiatiques sont mis au service de la néologie ». In *Actes des 7^{èmes} journées scientifiques du Réseau "Lexicologie, Terminologie, Traduction" (Agence Universitaire de la Francophonie) : « Mots, termes et contextes »*, Bruxelles, Belgique, pp.525-534.
- SAGER, J. C. (1990) *A Practical Course in Terminology Processing*. John Benjamins Publishing Company, Amsterdam/New York.
- SAND, A. & SIEMUND, R. (1992) « LOB – 30 Years on... ». *ICAME Journal*, 16, pp.119-122.
- SÁNDOR, Á. (2006) « Using the Author's Comments for Knowledge Discovery ». In *Actes de la Semaine de la Connaissance (SdC 2006)*, Nantes, France.
- SAUSSURE (DE), F. (1879) « Mémoire sur le système primitif des voyelles dans les langues indo-européennes ». *Revue Textes ! [en ligne]*, <http://www.revue-texto.net/index.php?id=1751>.
- SAUSSURE (DE), F. (1995) *Cours de Linguistique Générale*. (première édition 1916), Payot, Paris.

- SCHAETZEN (DE), C. (1989) *Terminologie diachronique*. Actes du colloque du 25-26 mars 1988, Cifl – Conseil International à la Langue Française, Bruxelles, Belgique.
- SCHIERZ, A. (2007) « Monitoring Knowledge: A Text-based Approach ». *Terminology*, 13(2), pp.125-154.
- SIMONDON, G. (2005) *L'invention dans les techniques*. Éditions du Seuil, Paris.
- SINCLAIR, J. (1991) *Corpus, Concordance, Collocation*. Describing English Language, Oxford University Press, Oxford.
- SLODZIAN, M. (2000) « L'émergence d'une terminologie textuelle et le retour du sens ». In H. Béjoint & P. Thoiron (Éd.), *Le sens en terminologie*, Presses Universitaires de Lyon, Travaux du CRTT (Centre de Recherche en Terminologie et Traduction), Lyon, pp.61-85.
- SOURNIA, J. C. (1994) « Les phases évolutives du vocabulaire médical français ». *Meta, Presses de l'Université de Montréal*, XXXIX(4), pp.692-700.
- STOJANOVIC, L. (2004) *Methods and Tools for Ontology Evolution*. Thèse de doctorat de l'Université de Karlsruhe, Allemagne.
- SUAREZ, M. & CABRÉ, M. T. (2002) « Terminological Variation in Specialized Texts: Linguistic Traces for Automatic Retrieval. ». In *Actes du VIII^{ème} Symposium ibéro-américain de terminologie (RITERM)*, Carthagène des Indes, Colombie, 28-31 octobre 2002.
- SWEETSER, E. E. (1990) *From Etymology to Pragmatics – Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge Studies in Linguistics, Cambridge University Press, Cambridge.
- TAAVITSAINEN, I. & PAHTA, P. (1997) « Corpus of Early English Medical Writing 1375 – 1750 ». *ICAME Journal*, 20, pp.71-78.
- TARTIER, A. (2003) « A Method for Observing Terminological Evolution ». In *Actes du colloque "Recent Advances in Natural Language Processing", (RANLP'03)*, G. Angelova, K. Bontcheva, R. Mitkov, N. Nicolov & N. Nikolov, Borovets, Bulgarie, pp.467-471.
- TARTIER, A. (2004) *Analyse automatique de l'évolution terminologique : variations et distances*. Thèse de doctorat en Informatique, Université de Nantes.

- TARTIER, A. (2006) « Variation terminologique et analyse diachronique ». In *Actes de la 13^{ème} conférence "Traitement Automatique des Langues Naturelles" (TALN 2006)*, Leuven, Belgique, 10-13 avril 2006, pp.347-356.
- TARTIER, A. (2007) « Analyse automatique de l'évolution terminologique ». In N. Aussenac-Gilles, A. Condamines & F. Sédès (Éd.), *Évolution et maintenance des ressources terminologiques*, Information – Interaction – Intelligence (I3), Hors série.
- TEMMERMAN, R. (2000a) *Towards New Ways of Terminological Description. The Sociocognitive Approach*. John Benjamins, Amsterdam/Philadelphie.
- TEMMERMAN, R. (2000b) « Une théorie réaliste de la terminologie : le socio-cognitivism ». *Terminologies Nouvelles*, 21, pp.58-64.
- THOIRON, P. & BÉJOINT, H. (2003) « La Terminologie, une question de termes ». In *Actes de la Conférence Internationale de Terminologie (CIT 2003)*, Lisbonne, Portugal, 11-13 décembre 2003.
- TISSAOUI, A. (2009) « Typologie de changements et leurs effets sur l'évolution de Ressources Terminologiques ». In *Actes de la conférence Ingénierie des Connaissances (IC 2009)*, F. Gandon, Hammamet, Tunisie.
- TOGNINI-BONELLI, E. (2001) *Corpus Linguistics at Work*. John Benjamins, Amsterdam/Philadelphia.
- TOUSSAINT, Y. (2004) « Extraction de connaissances à partir de textes structurés ». In J.-M. PINON & B. RUMPLER (Éd.), *Fouille de textes et organisation de documents*, Documents numériques, Hermès, 8(3), Paris, pp.11-34.
- TRAUGOTT, E. C. & DASHER, R. B. (2005) *Regularity in Semantic Change*. Cambridge Studies in Linguistics, Cambridge University Press, Cambridge.
- TURNER, P. (2004) « Word Sense Disambiguation by Web Mining for Word Co-Occurrence Probabilities ». In *Actes de la 3^{ème} conférence internationale "Evaluation of Systems for the Semantic Analysis of Text" (SENSEVAL-3 2004)*, Barcelone, Espagne, 25-26 juillet 2004.
- VALETTE, M. (2008) « Pour une science des textes instrumentée ». In M. Valette (Éd.), *Syntaxe et Sémantique – Textes, documents numériques, corpus. Pour une science des textes instrumentée*, Presses Universitaires de Caen, 9, Caen, pp.9-14.

- VALIN, R. (Éd.) (1990) *Les leçons de linguistique de Gustave Guillaume, 1948-1949. Grammaire particulière du français et grammaire générale (IV)*. Presses de l'Université de Laval, Québec.
- VAN CAMPENHOUDT, M. (1998) « Maille ou maillon : quand les terminographes négligent l'évolution de l'usage ». In A. Clas, S. Mejri & T. Baccouche (Éd.), *Actes des 5^{èmes} journées scientifiques du Réseau "Lexicologie, Terminologie, Traduction"* (Agence Universitaire de la Francophonie) : « La mémoire des mots », Tunis, Tunisie, pp.251-272.
- WESEMAËL, F. & WESEMAËL, R. (2005) « 1955-2005 : 50 ans de science reflétés dans le Petit Larousse Illustré ». In M. Cormier & A. Francoeur (Éd.), *Les dictionnaires Larousse. Genèse et Évolution*, Presses Universitaires de l'Université de Montréal, Collection Paramètres, Montréal, pp.199-217.
- WESTIN, I. & GEISLER, C. (2002) « A Multi-dimensional Study of Diachronic Variation in British Newspaper Editorials ». *ICAME Journal*, 26, pp.133-152.
- WÜSTER, E. (1981) « L'étude scientifique générale de la terminologie, zone frontalière entre la linguistique, la logique, l'ontologie, l'informatique et les sciences des choses ». In G. Rondeau & H. Felber (Éd.), *Textes choisis de terminologie*, GIRSTERM, Groupe Interdisciplinaire de Recherche Scientifique et Appliquée en Terminologie, I. Fondements théoriques de la terminologie, sous la direction de V.I. Siforov, Québec, Canada.
- YAROWSKY, D. (1995) « Unsupervised Word Sense Disambiguation Rivaling Supervised Methods ». In *Actes de la conférence "Association for Computational Linguistics" (ACL 1995)* (Éd.), Cambridge, MA, pp.189-196.
- ZWEIGENBAUM, P. & GRABAR, N. (2000) « Liens morphologiques et structuration de terminologie ». In *Actes de la conférence Ingénierie des Connaissances (IC 2000)*, Nantes, France, pp.325-334.
- ZWEIGENBAUM, P. & HABERT, B. (2004) « Accès mesurés au sens ». *Mots, Langue(s) et nationalisme(s)*, 74, pp.93-106.

Index thématique

A

Agence Spatiale Européenne..... 264
analyse distributionnelle 224–27
analyste (biais)..... 286
analyste (définition)..... 303
analyste (rôle)..... 303–5
AntConc (outil) 95
application 314–19

B

Barçah (logiciel)..... 192
Bateau de Thésée 269
bi-univocité..... 43
brachychrony..... Voir diachronie courte

C

centralité..... 212, 241, 242
Centre National d'Études Spatiales..... 2, 80
changement de type I..... 269
changement de type II..... 269
changement sémantique 12
Chi-Deux..... Voir χ^2
chrono-homogènes 109–10
chronologie externe 64
chronologie interne..... 64
Cnes .. Voir Centre National d'Études Spatiales
co-construction de l'interprétation (schéma)
..... 302
co-construction de l'interprétation 299–319
coefficient d'implantation..... 202
combinaison indices (évaluation)..... 292–98
concept 43
concept (définition) 50
concurrence (variantes)..... 201
connaissances (définition) 51, 53
consensus interprétatif..... 301
construction de l'interprétation 300
contexte de discours
..... Voir relation discursive contextuelle
contextes riches en connaissances évolutives
..... 148–87
cooccurrence (définition) 221–24
cooccurrence de surface 223

cooccurrence syntaxique 223
cooccurrence textuelle 223
corpus baromètres..... Voir corpus de suivi
corpus chronologiques
..... Voir corpus comparables
corpus comparables 69–70
corpus de suivi 67–69
corpus diachroniques (contraintes de
constitution 83
corpus ouverts Voir corpus de suivi

D

ddl Voir degrés de liberté
degrés de liberté (χ^2) 113
dépendances syntaxiques (définition)..... 221–24,
231
descendant en expansion 232
descendant en tête 232
diachronie (définition)..... 62
diachronie courte..... 17, 66
dimension externe 29, 41–60, 51
dimension interne 29, 41–60, 54
discours..... 54, 55, 267
domaine 45
domaine (définition)..... 46
domaine d'activité..... 46
domaine d'activité 79–82
domaine de connaissance..... 46, 79–82
DORIS (corpus)..... 88
Doris (projet)..... 88
double nature du terme..... 45

E

écart (χ^2) 113
empreintes de fréquence 106–7, 113
entretiens (formulation consigne) 287
ESA..... Voir Agence Spatiale Européenne
évolution en langue des signes..... 12
expert (définition) 305
expert (désaccords) 310
expert (qualités)..... 306
expert (rôle)..... 306
experts..... 78
experts (contraintes entretiens) 285–87

experts (exemples entretiens)..... 287–98
 extraction de termes 97–102

F

foisonnement néologique transitoire *Voir*
 foisonnement terminologique
 foisonnement terminologique..... 191, 200

G

genres textuels..... 12
 gestion des risques..... 152
 grammaire comparée 11
 grammaticalisation 12, 14

H

Histoire..... 57
 histoire de la langue 56
 histoire des sciences..... 57
 histoire des techniques..... 57
 histoire externe..... 13, 55
 histoire interne 13, 55

I

implantation terminologique 191
 indices linguistiques..... 73, 74
 interface (néologie) 229
 interprétation..... 51
 interprétation (construction) 51, 53

K

Khi-Deux..... *Voir* χ^2

L

linguistique "à l'instrument" 70–77
 linguistique diachronique..... 56
 linguistique historique 56
 linguistique outillée..... 70–77

M

marqueurs d'amélioration..... 164
 marqueurs d'ancrage dans le passé 161
 marqueurs d'ancrage dans le présent 160
 marqueurs d'évolution 148–87
 marqueurs d'insuffisance 162
 marqueurs d'obsolescence..... 162
 marqueurs de comparaison..... 163
 marqueurs de datation..... 159

marqueurs de nouveauté 161
 marqueurs de prédiction..... 165
 marqueurs de remplacement..... 163
 marqueurs de succession 163
 monitor corpus *Voir* corpus de suivi
 monoréférentialité..... 43

N

néologie 12, 107
 notion 43

O

obsolescence..... 108
 onomasiologie..... 49
 optique (éléments historiques) 80
 outils..... 70–77

P

paradigmes (Kuhn) 225
 parallèle entre évolution dans la langue et
 évolution des connaissances (hypothèse).....
 41–60
 périodisation des corpus 62–70
 philologie..... 11
 phonétique historique..... 11
 point de vue 244–49
 productivité morphologique et lexicale
 228, 238

R

réduction à caractère lexical *Voir* variante
 lexicale
 relation discursive contextuelle..... 170
 reprise anaphorique (variante).....
 *Voir* variante discursive
 réseau de dépendances..... 232
 ressources termino-ontologiques 52, 316
 ressources termino-ontologiques (maintenance
 et évolution)..... 316
 rétro-diagnostic 286, 309
 rétro-diagnostic (définition)..... 307
 RTO 316

S

sémantique cognitive 13
 sémantique structurale 12
 sémasiologie..... 49

sentiment d'évolution	307
sentiment néologique	285, 307, 308
série paradigmatique.....	232
seuil de fréquence	110
signifié normé.....	<i>Voir</i> concept (définition)
sociocognitivism	23
socioterminologie	23, 50, 55
stratégie "continuiste".....	13
synchronie (définition).....	62
synchronie dynamique	63, 178
synonymie de circonstance	201
Syntex (outil)	92–94, 231

T

terme (définition).....	42–48
terme étiquette	42
terme signe linguistique	44
terminochronie.....	33–35
terminologie computationnelle.....	35
terminologie diachronique	21–22
terminologie textuelle	25–27, 48–55
terminométrie.....	33
TerminoWeb (outil).....	96
TermOnto (outil).....	93

TermWatch (logiciel)	230
textes	54
TextTiling	168
TGT <i>Voir</i> Théorie Générale de la Terminologie
théorie de la dynamique des terminologies ..	24
Théorie Générale de la Terminologie	18, 42
TTVS.....	84
typologie de l'évolution.....	253–72

U

univocité	43
-----------------	----

V

variantes discursives.....	193–96, 214–19
variantes lexicales.....	193–96, 214–19
variantes (définition)	190
variantes (typologie).....	197–99
veille scientifique et technologique.....	150, 230

X

χ^2	111–15
----------------	--------

Index des auteurs

A

Achard, P	57
Acot, P	57, 265
Agbago, A	158
Ahmad, K.....	36, 64, 109, 228, 238
Aït El Mekki, T	314
Alber-Dewolf, R.....	193, 215
Al-Thubaity, A	36
Anthony, L.....	95
Antoine, G.....	56, 68
Audibert, L	222
Auger, P	21, 32
Auroux, S	71
Aussenac-Gilles, N	25, 27, 52, 79, 93, 148, 223, 300, 314, 317

B

Bachelard, G	299
Bachimont, B.....	25
Bae, H.S	326
Baneyx, A	93
Banks, D	21
Baroni, M.....	223
Barrière, C	96, 148, 158
Baudet, J.C	57
Beacco, J.C.....	245
Béciri, H.....	28
Béjoint, H	42, 43, 49, 55, 227, 306
Beltran-Vidal, D	21
Ben Hariz Ouenniche, S.....	285, 308
Benvéniste, É	54
Bertels, A	222
Bessé (de), B.....	46, 80, 82
Biber, D.....	62, 71
Blackwell, S.....	68
Blank, A.....	12, 13
Blumenthal, P	222
Bonhomme, M.....	194
Bonnal-Vergès, F	12
Borillo, A	148, 169, 170
Bouaud, J.....	223
Boulanger, J.C.....	20, 28

Bourigault, D.....	3, 25, 47, 48, 52, 68, 79, 92, 93, 100, 223, 231, 233, 300, 303, 306, 317
Boutayeb, S.....	191
Boutet, J.....	300, 304
Bouveret, M.....	43
Bowker, L	31, 62, 68, 69, 85, 107
Bréal, M.....	10
Brunot, F	56
Buitelaar, P	317

C

Cabré, M.T.....	31, 44, 46, 59, 78, 107, 148
Calberg-Challot, M.....	69, 310
Candel, D.....	18
Carreño, I.....	97
Cartier, É.....	31, 68
Celotti, N	2
Château, C.....	64
Church, K	225
Cimino, J.J.....	316
Collet, T	43, 193, 215, 267
Collier, A	68
Condamines, A	25, 43, 45, 47, 50, 71, 80, 89, 92, 148, 157, 169, 227, 244, 268, 300, 304, 306, 314
Conein, B	57
Coseriu, E	12
Courtois, B.....	68
Croft, W	11
Cusin-Berche, F.....	268

D

Daille, B.....	196, 203, 204
Darmesteter, A.....	10
Dasher, R. B.....	11
De Schaetzen (de), C	20
De Surmont, J.N.....	67
Dejean, H.....	69
Delaporte, Y	12
Delavigne, V.....	21, 46, 47, 97
Delesalle, S	10
Depecker, L	32, 43, 45, 46, 51, 192
Desmet, I.....	29, 43
Devereaux Ferguson, S.....	287

Drouin, P31, 64, 65, 66, 78, 107, 108, 111
Dury, P
1, 19, 20, 21, 22, 43, 47, 49, 60, 66, 70, 79, 108,
109, 191, 201, 262, 286, 307

E

Eckardt, R..... 12
Engwall, G 70
Estopà Bajot, R..... 79, 310
Evert, S 223

F

Fabre, C 93, 222, 225
Fairon, C..... 68
Faulstich, E..... 190
Felber, H..... 18, 43
Ferret, O 203
Ferret, S 265, 268
Feuillard, C 63
Fiala, P 57, 67
Firth, J.R 221
Freixa, J..... 190
Fuchs, C..... 76

G

Galarreta, D 3
Galy, É 68
Gardin, B 285, 307
Gargouri, Y 317
Gasiglia, N 71
Gaudin, F 23, 32, 43, 44, 45, 50, 55, 63
Gaussier, É 69
Geeraerts, D 13
Geisler, C..... 12, 263
Gentilhomme, Y 43
Gillam, L..... 317
Gouadec, D 305
Grabar, N 196
Grefenstette, G 222, 224
Groult, M..... 21
Grzega, J 108
Guespin, L..... 1, 23, 24
Guilbert, L..... 23, 28, 29, 57, 59, 63, 191, 195, 200

H

Habert, B.....
62, 68, 69, 70, 71, 76, 107, 109, 132, 223, 225,
300, 326
Hamelin, L.E 28, 49
Hamon, T..... 148, 313
Harris, Z.S 222, 224–27, 300
Haton, J.P..... 53
Hearst, M..... 148, 168, 222
Heiden, S 15
Helgorsky, F..... 41, 56, 64, 65
Ho-Dac, L.M..... 70, 73, 187
Hoenigswald, H 11
Holmes, J 14, 16, 69
Houdebine, A.M..... 63
Howell, D. C..... 111, 113
Humbley, J 1, 19, 28, 29, 49, 107, 229, 239
Hundt, M..... 14

I

Ibekwe-Sanjuan, F 149, 168, 191, 230
Iezzi, D. F..... 12

J

Jacquart, H..... 287
Jacquemin, C 52, 196, 203
Jacques, M.P 193, 209, 215, 267
Janssen, M 31, 107
Josselin-Leray, A..... 19, 42, 43, 46, 49, 287

K

Kageura, K..... 24, 25
Karlgren, J..... 112
Kehoe, A 14
Kilgarriff, A 222
Klein, M 316
Koch, P 12, 13
Kocourek, R..... 21, 48, 54, 193
Krier, F 63
Kristeva, J 85
Kuhn, T 64, 225
Kytö, M 17, 70

L

L'Homme, M.C
..... 43, 44, 47, 51, 52, 69, 93, 97, 148, 250, 326

Laignelet, M..... 154, 168, 186
 Lamé, G 79, 223, 303
 Lass, R 11, 73
 Lavrentiev, A..... 15
 Lebel, M.É 12, 73
 Leblanc, J.M 227
 Leech, G..... 12, 65
 Lelubre, X..... 80
 Lerat, P 45
 Leroy, S..... 71, 158
 Lervad, S 19, 66, 191
 Lino, M. T..... 28
 Losee, R.M 106, 222
 Lozachmeur, G 263
 Luong, P 316

M

Mackintosh, K 19, 21
 Maedche, A..... 317
 Mair, C..... 12, 14, 17, 65, 66
 Malaisé, V 171
 Maniez, F..... 21
 Marchello-Nizia, C 12, 14, 15, 55, 56, 64, 73
 Marshman, E 148, 157
 Martinet, A..... 63
 Martinez, W 227
 Mathieu, Y 30, 107, 108
 Mayaffre, D..... 71
 McEnery, T..... 62, 71
 Meillet, A..... 10, 13
 Mel'čuk, I.A 326
 Méla, A 71, 158
 Ménard, N..... 111
 Meurman-Solin, A 14, 16, 70
 Meyer, I 19, 21, 148, 157
 Minett, J 31
 Moirand, S 245
 Møller, B..... 1, 20, 33-35, 60, 77, 263
 Morin, E..... 203
 Mortureux, M.F..... 268
 Muller, C 111, 113
 Musacchio, M.T..... 2, 36, 64, 109

N

Nazarenko, A 223, 226, 313, 314
 Nicolas, A..... 10
 Normand, C..... 10

Nyckees, V.....13
 Nyrop, K.....29

O

Oliver, D316
 Ollinger, S.....107
 Otman, G25, 148, 305, 307
 Ottens, K.....313, 317, 319

P

Pahta, P14, 16
 Paquin, A.....31, 78, 107
 Pearson, J62, 68, 69, 85
 Péry-Woodley, M.P62, 71, 186
 Petit, G42, 45, 54
 Picton, A1
 Pimm, C168
 Poibeau, T46
 Portelance, C195
 Prévost, S..... 14, 15, 65, 67, 79
 Prost, A57
 Pruvost, J267

Q

Quemada, B.....21, 193
 Quirion, J32, 33, 192, 202

R

Rastier, F12, 13, 25, 268, 299
 Rebeyrolle, J 25, 43, 47, 89, 92, 227, 244
 Renouf, A.....14, 68, 73
 Rey, A.....2, 12, 28, 43, 44
 Reymond, D222
 Roche, S.....31, 107
 Rogozan, D316
 Rondeau, G.....18, 28, 43, 305
 Rothenburger, B.....3
 Ryckman, T224, 225

S

Sablayrolles, J.F.. 12, 28, 29, 31, 68, 267, 285, 308
 Sader Feghali, L107
 Sager, J.C.....1, 20, 46, 49
 Sand, A.....14, 16
 Sándor, Á.....152
 Sanjuan, E191, 230
 Saussure (de), F.....10, 11, 13, 41, 62, 65

Schierz, A 36, 109
 Séguéla, P 148
 Siemund, R 14, 16
 Simondon, G..... 266
 Sinclair, J 68
 Slodzian, M..... 3, 25, 46, 47, 48, 92
 Soubeille, A..... 89
 Sournia, J. C 21
 Staab, S 317
 Stojanovic, L 316, 317
 Suarez, M 148
 Sweetser, E. E 12, 13
 Szulman, S 314

T

Taavitsainen, I 14, 16
 Tartier, A 36, 37, 190
 Temmerman, R..... 23
 Thoiron, P 42, 43, 49, 306
 Tissaoui, A 316, 317
 Tognini-Bonelli, E 71
 Toussaint, Y 186
 Traugott, E. C 11

Turney, P 222

V

Valette, M 72, 107
 Valin, R 11
 Van Campenhoudt, M 1, 21, 49
 Véronis, J 68

W

Wang, W.S.Y 31
 Wesemaël, F 77
 Wesemaël, R..... 77
 Westin, I..... 12, 263
 Wilson, A..... 62, 71
 Wüster, E 18, 19, 24

Y

Yarowsky, D..... 222
 Yzagurri, L 31

Z

Zweigenbaum, P 196, 225, 300, 326

Index des termes/concepts traités

A

<i>aberrations</i>	135
<i>absorption</i>	135
<i>acquittements</i>	126, 128
<i>activation satellite</i>	240
<i>adhérence</i>	115
<i>afficheur</i>	125, 126, 129, 140, 281–84
<i>afficheur indiquer</i>	125, 126, 130, 140
<i>alignement</i>	135
<i>altitude de satellite</i>	240
<i>Ampli</i>	216
<i>amplificateur</i>	216
<i>anomalie</i>	213, 256
<i>antenne</i>	135, 279–81
<i>appuyer</i>	125, 126, 130, 140
<i>appuyer sur touche</i>	125, 126, 130, 140
<i>APS</i>	115, 117, 118, 135, 254
<i>arêtes</i>	115, 117, 119, 135, 184
<i>ASIC</i>	213
<i>astigmatisme</i>	135
<i>autorisation de satellite</i>	240
<i>autoriser</i>	140
<i>axe</i>	135

B

<i>baffle</i>	135
<i>balise maîtresse</i>	212
<i>balise maîtresse DORIS</i>	212
<i>bi-CMOS</i>	212, 276–79
<i>BiCMOS</i>	182, 212, 258, 276–79
<i>bloc</i>	140, 281–84
<i>bloc de télégestion</i>	209
<i>bloc froid</i>	279–81
<i>bloc TLG</i>	209
<i>blocs</i>	126, 127
<i>bruit</i>	135, 279–81

C

<i>cavité</i>	135
<i>cavité étendue</i>	117, 120, 135
<i>CCD</i>	126, 128
<i>CCI</i>	126, 128
<i>CDC</i>	128

<i>centre de contrôle DORIS</i>	128
<i>circuit</i>	238
<i>circuit hybride</i>	238
<i>clavier</i>	140, 281–84
<i>clé</i>	140
<i>CMOS</i>	182, 276–79
<i>cœur</i>	117, 135
<i>coffret batterie</i>	212, 255
<i>coffret de batterie</i>	212, 255
<i>colle</i>	117, 120, 135, 261
<i>configuration</i>	126, 127, 140, 281–84
<i>constellations de satellites</i>	183
<i>contexte</i>	126, 127, 140, 281–84
<i>coordonnées de satellite</i>	240
<i>correcteur</i>	242
<i>couches</i>	135
<i>couches minces optiques</i>	184
<i>courbure</i>	135
<i>créneau</i>	126
<i>cryostat</i>	279–81

D

<i>DCM</i>	209
<i>Décentrement Miroir Concave</i>	209
<i>déformation</i>	135
<i>détecteur infrarouge</i>	181
<i>détecteur thermique</i>	181
<i>détecteur pneumatique</i>	255
<i>diffusion</i>	135
<i>dimensionnement</i>	135
<i>dimensionnement radiométrique</i>	279–81
<i>diode</i>	120
<i>diodes lasers</i>	179, 180
<i>domaine spectral</i>	135
<i>donnée</i>	140, 281–84

E

<i>élément parasite</i>	246
<i>émission</i>	281–84
<i>émission des blocs de téléchargement</i>	209
<i>émission des TCHs</i>	209
<i>espace froid</i>	135, 181, 279–81
<i>étalonnage</i>	260

F

<i>face</i>	135
<i>facteur de remplissage</i>	115
<i>fibres</i>	135
<i>fibres dopées</i>	115
<i>fibres monomodes</i>	179, 180
<i>fibres multimodes</i>	180
<i>fibres optiques</i>	119, 179, 180, 208
<i>fibres optiques dopées</i>	208
<i>fichiers</i>	126, 127, 140, 258, 281–84
<i>flux lumineux parasite</i>	218, 246
<i>flux parasite</i>	184, 218, 246
<i>fonction</i>	239
<i>fonction recalage en temps</i>	239, 258
<i>froid</i>	279–81

G

<i>GEO</i>	135
<i>gestion</i>	140
<i>GOMOS (projet)</i>	242

H

<i>Hélium</i>	279–81
<i>HRG</i>	117, 118

I

<i>IASI (projet)</i>	212
<i>identification</i>	140
<i>identification de satellites</i>	240
<i>IHM</i>	124, 258, 281–84
<i>image</i>	135
<i>image parasite</i>	184, 246
<i>imageurs</i>	123
<i>Infrarouge</i>	211, 216, 259, 281
<i>Infra-rouge</i>	211
<i>Infrarouge Thermique</i>	211, 216, 259
<i>instrument imageur</i>	117, 123
<i>instrumental</i>	135
<i>instruments imageurs</i>	116
<i>interface</i>	140, 245, 281–84
<i>interface CCD</i>	281–84
<i>interface entre capteurs</i>	281–84
<i>interface opérateur</i>	245, 281–84
<i>interface télégestion</i>	281–84
<i>interface vers le CDC</i>	281–84
<i>interrupteur secteur</i>	256, 262

<i>IR</i>	211
<i>IRT</i>	211

L

<i>lampes flash</i>	180
<i>laser semi-conducteur monomode</i>	233
<i>laser à cavité étendue</i>	233
<i>laser à fibres dopées</i>	233
<i>laser à puits quantiques</i>	233
<i>laser accordable</i>	233
<i>laser de communication</i>	233
<i>Laser Diode Transmitter Package</i>	208
<i>laser monomode</i>	233
<i>laser pompé par diode</i>	233
<i>laser semi-conducteur</i>	233
<i>laser solide</i>	233
<i>lasers</i>	119, 120, 179, 234
<i>LDEF (projet)</i>	117, 123
<i>LDTP</i>	208
<i>LEO</i>	135
<i>liaison entre satellites</i>	210, 211, 217
<i>liaison intersatellite</i>	179, 183, 210, 211, 217
<i>liaison inter-satellites</i>	210
<i>liaisons LEO-GEO</i>	211
<i>logiciel</i>	126, 127, 140, 281–84
<i>LSB</i>	126, 128
<i>lumière parasite</i>	246, 248
<i>luminance</i>	135, 279–81

M

<i>machine cryogénique</i>	181, 184
<i>machine cryogénique active</i>	181
<i>microbolomètres</i>	181
<i>microns</i>	115
<i>milieu</i>	135
<i>minisatellite</i>	207
<i>mini-satellite</i>	207, 238, 254
<i>MIR</i>	211
<i>microbolomètres</i>	279–81
<i>miroir</i>	135, 136, 178, 184, 241, 242
<i>miroir asphérique</i>	184
<i>miroir concave</i>	242
<i>miroir concave correcteur</i>	241
<i>miroir concave correcteur arrière</i>	241
<i>miroir hors d'axe</i>	241
<i>miroir mère</i>	241
<i>miroir plan</i>	184

<i>miroir primaire</i>	184
<i>miroir secondaire</i>	241
<i>miroir sélectif</i>	241
<i>miroir sphérique</i>	241
<i>Modèle Instrumental</i>	116, 117, 122, 129, 257
<i>Modèle Radiométrique Instrumental</i>	116, 117, 122
<i>module de gestion</i>	126, 127, 140, 281–84
<i>MOS</i>	135, 276–79
<i>Moyen Infrarouge</i>	211, 216, 259
<i>Moyen-Infrarouge</i>	211
<i>multi-barrettes</i>	178

N

<i>non refroidi</i>	181, 279–81
---------------------------	-------------

O

<i>opérateur</i>	140, 245, 281–84
<i>opérateur appuyer</i>	126, 130, 140, 245
<i>opérateur appuyer sur touche</i>	126, 130, 140, 245, 281–84
<i>opérateur appuyer sur val</i>	245, 281–84
<i>opérateur tourner clé</i>	245, 281–84
<i>optique</i>	135
<i>orbite de satellite</i>	240
<i>orbite géostationnaire</i>	115
<i>oscillateur ultra-stable</i>	212, 258
<i>OUS</i>	212

P

<i>palier de référence</i>	212
<i>palier référence</i>	212
<i>panne</i>	213, 256
<i>PAO</i>	208
<i>parasites</i>	135, 246
<i>passage</i>	140
<i>passage de satellite</i>	240
<i>Pharao (projet)</i>	120, 242
<i>photo</i>	135
<i>PIR</i>	211
<i>pixel</i>	135
<i>plan</i>	135
<i>polissage</i>	113, 115, 135, 184, 233, 242
<i>polissage assisté par ordinateur</i>	208, 233, 242
<i>polissage de dioptré</i>	233, 242
<i>polissage de lentille</i>	233, 242
<i>polynômes</i>	183
<i>pompage optique</i>	180

<i>position</i>	140
<i>Proche Infrarouge</i>	211, 216, 259
<i>Proche-Infrarouge</i>	211
<i>puce détectrice</i>	178

R

<i>radiateur</i>	184
<i>radiateur passif</i>	181
<i>radiomètres imageurs</i>	123
<i>radiométrique</i>	135
<i>raie</i>	135
<i>rapport S/B</i>	201
<i>rapport signal à bruit</i>	201
<i>rapport signal sur bruit</i>	201
<i>rapport signal/bruit</i>	201
<i>rapport signal-sur-bruit</i>	201
<i>rayonnement parasite</i>	246
<i>réception</i>	140, 281–84
<i>recevoir</i>	140, 281–84
<i>reçus</i>	126, 127
<i>refroidi à l'Hélium</i>	279–81
<i>refroidi par de l'Hélium</i>	279–81
<i>refroidir</i>	115, 135, 279–81
<i>refroidir l'optique</i>	279–81
<i>refroidir l'instrument</i>	279–81
<i>refroidissement</i>	135, 181, 279–81
<i>refroidissement passif</i>	279–81

S

<i>S/B</i>	201
<i>satellite</i>	140, 238, 240
<i>signal</i>	135, 279–81
<i>signal parasite</i>	246
<i>Silex (projet)</i>	179
<i>SO</i>	210
<i>sondeur</i>	242
<i>sondeur au limbe</i>	242
<i>sondeur passif</i>	242
<i>sondeur vertical</i>	242
<i>source laser de communication</i>	208
<i>source parasite</i>	246
<i>spécifications d'interface</i>	281–84
<i>Spot (projet)</i>	118, 122, 183
<i>synthèse d'ouverture</i>	123, 210, 256
<i>synthèse d'ouverture optique</i>	116, 117, 123, 184, 210, 256
<i>système optique</i>	115

T

TCH.....	124, 126, 127, 140, 281-84
TDI.....	135
téléchargement.....	258, 281-84
télécommunications intersatellites.....	179
télégestion.....	124, 126, 127, 140, 281-84
télégestion logicielle.....	209
terminal de transfert.....	124, 126, 128
TLG.....	126, 127, 140
TLGL.....	209
TLGP.....	126, 127, 140, 281-84
touche.....	125, 126, 129, 140, 239
touche eff.....	239
touche hrl.....	239
touche rct.....	239
touche seq.....	239
touche syn.....	239
touche tst.....	239

<i>touche val</i>	126, 129, 140, 239, 256
<i>trajet parasite</i>	246
<i>transfert des charges</i>	135
<i>transition</i>	117, 121
<i>transition optique</i>	121
<i>transition vitreuse</i>	121

U

<i>usinage ionique</i>	178, 184
------------------------------	----------

V

VAL.....	126, 140
vecteur de satellite.....	240
VIS.....	211
visibilité.....	124, 126, 140
visibilité de satellite.....	201, 208, 240
visibilité satellite.....	127, 201, 208, 240
visible.....	211
Visible.....	216

Annexes

ANNEXE A – Extraits des corpus

Corpus TTVS

5.3.2. ARCHITECTURE D'UNE LIAISON OPTIQUE

5.3.2.1. DESCRIPTION GENERALE

Le système de communication s'effectue, soit entre deux terminaux situés sur deux satellites différents, soit entre un satellite et une station de réception au sol.

Dans le cas d'une liaison entre satellites, il existe encore deux classes d'applications : le relais entre satellites géostationnaires situés à grande distance l'un de l'autre et échangeant de très hauts flots de données (typiquement 1 Gigabits-1) et le satellite en orbite basse fournissant les données à haut débit vers un relais géostationnaire.

La liaison satellite-sol n'est utilisée que dans des cas exceptionnels, car la transmission, la turbulence et la diffusion atmosphérique perturbent et ne permettent pas d'assurer la fonction, à cent pour cent du temps.

On peut citer, à titre d'exemple, les liaisons entre sondes lointaines (exploration des planètes proches) et la Terre. Le débit de photons étant extrêmement faible, à cause des très grandes distances (supérieur à 1 à 5 Unité Astronomique), il est nécessaire d'augmenter la surface collectrice des antennes. Des études récentes (CNES, JPL...) montrent que les diamètres des télescopes collecteurs doivent être de l'ordre de 2 à 4 m. On conçoit qu'il est beaucoup plus aisé d'implanter ces antennes donc des sites bien choisis sur Terre, que de développer un satellite en orbite basse accueillant une telle antenne.

[...]

Ce dépointage grossier est assuré par un mécanisme à deux axes de rotation indépendants ; l'opération d'alignement sommaire s'effectue généralement juste avant la transmission, quand les deux satellites ne sont pas encore en vue directe, grâce aux informations connues (éphémérides des deux satellites et connaissances "a priori" de l'attitude de chacun des deux satellites).

La précision d'orientation du mécanisme grossier est limitée par la connaissance de l'attitude du satellite, soit typiquement quelques dixièmes de degrés.

Un délai de quelques minutes est généralement nécessaire pour prépositionner correctement l'antenne optique, sans perturber l'attitude du satellite.

La réception s'effectue symétriquement par rapport à l'émission : on retrouve sur le deuxième terminal, les mêmes fonctions.

Notons qu'en plus, sur le terminal géostationnaire (GEO) (ou sur l'un d'entre eux), il existe une balise lumineuse destinée au repérage de ce terminal par l'autre satellite.

Emettant, dans un angle solide correspondant au cône d'incertitude d'acquisition, un flux lumineux, continu ou légèrement modulé, cette balise permet au terminal en orbite basse (LEO) de repérer la position angulaire du satellite géostationnaire (GEO) et de pointer dessus, son axe optique.

[...]

Soulignons enfin, pour terminer cette description sommaire du système, la présence de filtres optiques qui permettent :

- de sélectionner le signal et de rejeter le maximum de lumière parasite,
- de multiplexer et de démultiplexer les flux lumineux. Ceci veut dire, qu'on place sur le même trajet lumineux, plusieurs voies d'émission #cs#1, #cs#2... #cs#n à des longueurs d'onde légèrement différentes. On effectue l'opération symétrique sur l'autre terminal (démultiplexeur). Cette opération

permet d'augmenter la puissance lumineuse émise, en multipliant le nombre de lasers aux longueurs d'onde respectives $\lambda_1, \lambda_2 \dots \lambda_n$,

- de coder la lumière en jouant sur les états de polarisation : on peut émettre une lumière polarisée rectilignement, circulaire droite, gauche, etc...

5.3.2.2. SOURCES LASER-MODULATION-DETECTION

5.3.2.2.1. LES SOURCES LASER

Les principales exigences spatiales pour la source lumineuse peuvent être résumées comme suit :

- modulable aisément à de très hautes cadences (typiquement quelques centaines de mégahertz),
 - de grande fiabilité : on souhaite un fonctionnement sans défaillance pendant plusieurs dizaines de milliers d'heures,
 - une très grande brillance : la source doit émettre la puissance maximum dans l'angle solide minimal ; cette source doit donc être monomode transverse ; si on désire travailler en détection cohérente (homodyne ou hétérodyne), il faut disposer en outre d'une source monomode longitudinale ;
 - un bon rendement (puissance lumineuse par rapport à la puissance électrique fournie),
 - le minimum de contraintes opérationnelles et d'embarquabilité : le laser ne doit pas être trop lourd ni encombrant. Il doit fonctionner, de préférence, à température ambiante sans nécessiter un besoin de stabilisation trop sévère. Il doit, en outre, être naturellement suffisamment robuste pour résister aux ambiances vibratoires mécaniques et acoustiques du lancement.
-

Corpus DORIS

4. DESCRIPTION DE LA BALISE 3.0

4.1. SYNOPTIQUE

Le synoptique suivant présente l'ensemble balise dans son environnement électrique.

Figure 11 : Synoptique de la balise dans son environnement

4.2. ELEMENTS CONSTITUTIFS DE LA BALISE

La dénomination de l'ensemble balise recouvre tout un ensemble de matériel :

- Le coffret d'émission et de gestion comprenant :
- le synthétiseur bi-fréquence,
- le module amplification,
- le module de gestion,
- le module OUS (comprenant l'OUS et la carte de consigne),
- l'antenne bi-fréquence,
- la station météo,
- les câbles reliant les différents éléments de la balise,
- le coffret d'alimentation.

4.3. FONCTIONNEMENT

L'installation de l'ensemble balise consiste à connecter au coffret d'émission et de gestion :

[...]

5. SPECIFICATIONS TECHNIQUES DU COFFRET D'EMISSION ET DE GESTION

5.1. LE MODULE OUS

Les chaînes de fréquences sont asservies sur des OUS dont les spécifications sont décrites dans le document DA2.

Le module OUS fournit à la balise la fréquence d'horloge de référence. Cette fréquence permettra :

- d'asservir le synthétiseur,
- de piloter le générateur du module de gestion.

[...]

5.4.5.2. Le mode 'PROGRAMMATION'

Description du mode : Ce mode permet de configurer la balise qui enregistrera les paramètres nécessaires pour qu'elle puisse remplir sa mission.

Tous les paramètres de programmation doivent être stockés dans une mémoire EEPROM, de manière à conserver la configuration de la balise en cas de coupure d'alimentation de la balise (ceci afin d'assurer un redémarrage automatique de cette dernière).

Le mode doit permettre de calculer les prévisions de passage des satellites autorisés sur les 24 heures suivant la date courante.

Sortie du mode : La balise reste dans le mode programmation tant que l'opérateur n'a pas transmis un ordre de changement de mode. En sortie de ce mode programmation, la balise passe en mode 'EMISSION', en mode 'AUTOTEST' ou en mode 'SURVIE'.

5.4.5.3. Le mode 'EMISSION'

Description du mode : Ce mode permet à la balise d'émettre les signaux DORIS sur les deux voies.

Les exigences liées à ce mode sont les suivantes :

- Emission au maximum 10 ms après le début de l'intervalle de temps pendant lequel elle doit émettre.
- Pas de saut de phase ni de coupure dans le cas d'une émission continue non séquencée,

- Pas d'émission si le synthétiseur est déverrouillé,
- Pas d'émission si la séquence n'est pas autorisée,
- Economie d'énergie sur une séquence sans émission.

Pour chaque créneau d'émission (de durée 10s), la balise doit donc vérifier si la séquence est autorisée ou non. La condition d'autorisation doit satisfaire l'ensemble des critères suivants :

- Digit du mot minute = 1,
- Digit du mot semaine = 1,
- satellite autorisé visible et satellite interdit non visible, ou bien émission programmé continue (séquencée par les mots semaine et minute),
- Fonction veille non activée,

Le mode 'EMISSION' peut être de type normal ou spécifique.

Description du mode en fonctionnement normal : Ce mode permet à la balise de remplir sa mission de localisation ou d'orbitographie. Seul ce mode l'autorise à émettre vers le satellite selon les critères définis lors de la programmation. La balise est alors entièrement autonome et elle émet vers le satellite selon le séquençement programmé.

Dans ce mode, le module de gestion crée le message Doris à émettre.

Description du mode en fonctionnement spécifique : Ce type de fonctionnement permet au CCD de recevoir la balise après un démarrage sans mise à l'heure ou après un redémarrage automatique avec perte de l'heure. Cette émission spécifique est indiquée dans le message émis par la balise. La suppression de ce mode d'émission spécifique doit être confirmée par IHM ou télégestion après mise à l'heure et synchronisation.

ANNEXE B – Précision des marqueurs d'évolution (TTVS)

		1994			1998			2002			
		Occurrences totales	Perti- nents	Précision TTVS1994	Occurrences totales	Perti- nents	Précision TTVS1998	Occurrences totales	Perti- nents	Précision TTVS2002	Précision globale
Marqueurs de datation	Marqueurs de datation										
	<i>années [date]</i>	1	1	100	3	3	100	2	2	100	100
	<i>en [date]</i>	0	0	nc.	0	0	nc.	2	2	100	100
	<i>depuis quelques années</i>	0	0	nc.	1	1	100	2	2	100	100
	Marqueurs d'ancrage dans le présent/l'actualité										
	<i>à ce jour</i>	0	0	nc.	1	1	100	2	2	100	100
	<i>actuel</i>	5	5	100	7	7	100	16	16	100	100
	<i>aujourd'hui</i>	1	1	100	4	4	100	4	4	100	100
	<i>classique</i>	0	0	nc.	1	1	100	9	6	66,67	70
	<i>disponible</i>	0	0	nc.	4	3	75	5	3	60	66,67
	<i>exister</i>	0	0	nc.	3	2	66,67	5	3	60	62,5
	<i>maintenant</i>	0	0	nc.	5	5	100	3	2	66,67	87,5
	Marqueurs d'ancrage dans le passé										
	<i>autrefois</i>	0	0	nc.	1	1	100	0	0	nc.	100
	<i>histoire</i>	0	0	nc.	1	1	100	1	1	100	100

Nouveauté et disparition de termes/concepts	<i>passé</i>	0	0	nc.	0	0	nc.	2	2	100	100
	Marqueurs de nouveauté/apparition										
	<i>apparaître</i>	3	2	66,67	7	6	85,71	8	7	87,5	83,33
	<i>arrivée</i>	1	0	0	2	1	50	1	1	100	50
	<i>inventer</i>	0	0	nc.	1	1	100	0	0	nc.	100
	<i>nouveau</i>	8	7	87,5	10	9	90	15	15	100	93,94
	<i>récent</i>	1	1	100	1	1	100	9	9	100	100
	Marqueurs du passé, disparition, obsolescence										
	<i>abandonner</i>	0	0	nc.	1	1	100	1	1	100	100
	Marqueurs d'insuffisance										
Nouveaux besoins potentiels	<i>malheureu- sement</i>	2	2	100	2	2	100	3	3	100	100
	<i>pas assez</i>	0	0	nc.	1	0	0	1	1	100	50
	<i>pas encore</i>	4	4	100	5	5	100	4	4	100	100
	<i>pas satisfaisant</i>	0	0	nc.	1	1	100	1	1	100	100
	Marqueurs de comparaison										
	<i>meilleur</i>	6	6	100	7	3	42,86	11	7	63,64	66,67
	<i>moins [...] que</i>	0	0	nc.	2	2	100	7	3	42,86	55,56
	<i>plus [...] que</i>	9	9	100	36	16	44,44	31	23	74,19	63,16
	Marqueurs de remplacement/succession										
	<i>remplacer</i>	0	0	nc.	3	3	100	2	2	100	100
	<i>successeur</i>	2	2	100	1	1	100	4	4	100	100
	Marqueurs d'amélioration/développement										
	<i>être à l'étude</i>	0	0	nc.	0	0	nc.	1	1	100	100

<i>amélioration / améliorer</i>	1	1	100	18	8	44,44	16	10	62,5	54,29
<i>développement / développer</i>	3	2	66,67	7	4	57,14	12	11	91,67	77,27
<i>encouragement</i>	1	1	100	0	0	nc.	0	0	nc.	100
<i>essor</i>	0	0	nc.	1	1	100	1	1	100	100
<i>évoluer/évolution / évolutif</i>	4	3	75	9	7	77,78	10	8	80	78,26
<i>progrès</i>	3	3	100	4	4	100	5	5	100	100
<i>prototype</i>	1	1	100	1	1	100	0	0	nc.	100
<i>réduire</i>	7	1	14,29	15	8	53,33	21	14	66,67	53,49
Futur										
<i>avenir</i>	3	3	100	2	2	100	1	1	100	100
<i>envisager</i>	3	3	100	5	4	80	5	4	80	84,62
<i>futur</i>	1	1	100	1	1	100	5	5	100	100
<i>imaginer</i>	1	1	100	2	2	100	2	2	100	100
<i>prometteur</i>	0	0	nc.	1	1	100	1	1	100	100
TOTAL / MOYENNE	71	60	84,51	177	124	70,06	231	189	81,82	77,87

ANNEXE C – Typologie de l'évolution en diachronie courte

	Interprétations			
	Aspect "extralinguistique"	Aspect "linguistique" en corpus	Exemples	Indices possibles
1	Apparition d'un nouveau concept dans le domaine	Apparition d'un terme/concept dans la terminologie du corpus (néologie de forme et de sens)	<i>APS, mini-satellite</i> (TTVS)	<ul style="list-style-type: none"> - fréquence (rupture) - contextes - variantes - dépendances (productivité)
2	Obsolescence d'un concept dans le domaine	Disparition d'un terme/concept obsolète dans la terminologie du corpus	<i>détecteur pneumatique</i> (TTVS)	<ul style="list-style-type: none"> - fréquence (rupture) - contextes - dépendances
3	Apparition non planifiée d'une variante	Apparition d'une variante en corpus	<i>coffret de batterie</i> → <i>coffret batterie</i> (DORIS)	<ul style="list-style-type: none"> - variantes - fréquence (rupture)
4	Disparition non planifiée d'une variante	Disparition d'une variante en corpus	<i>synthèse d'ouverture optique</i> → <i>synthèse d'ouverture</i> (TTVS)	<ul style="list-style-type: none"> - variantes - fréquence (rupture)
5	Apparition planifiée d'une variante	Apparition d'une variante en corpus	<i>anomalie</i> (DORIS)	<ul style="list-style-type: none"> - variantes - fréquence (rupture)
6	Disparition planifiée d'une variante	Disparition d'une variante en corpus	<i>panne</i> (DORIS)	<ul style="list-style-type: none"> - variantes - fréquence (rupture)
7	Apparition d'une instance	Apparition d'un terme/concept dans la terminologie du corpus	pas d'exemples attestés en corpus	
8	Disparition d'une instance	Disparition d'un terme/concept dans la terminologie du corpus	<i>interrupteur secteur</i> (DORIS)	<ul style="list-style-type: none"> - fréquence (formes chrono-

			<i>Modèle Instrumental</i> (TTVS)	homogènes)
9	Apparition de fonctionnalités	Apparition de termes/concepts dans un domaine d'activité et qui ne sont pas des néologismes	<i>visibilité satellite</i> (DORIS)	<ul style="list-style-type: none"> - fréquence (rupture ou formes chrono-homogènes) - variantes - contextes - dépendances
10	Disparition de fonctionnalités	Disparition de termes/concepts dans un domaine d'activité	<i>fonction recalage en temps</i> (DORIS)	<ul style="list-style-type: none"> - fréquence (rupture ou formes chrono-homogènes) - variantes - contextes - dépendances
11	Stabilisation/implantation d'un concept (familiarisation)	Implantation d'une variante (après une période de concurrence)	<i>Bi-CMOS</i> → <i>BiCMOS</i> (TTVS) <i>oscillateur ultra-stable</i> → <i>OUS</i> (DORIS)	<ul style="list-style-type: none"> - variantes - fréquence (rupture)
12	Migration de termes/concepts d'un domaine connexe	Apparition de termes dans le corpus Changement de point de vue	Termes de l'informatique (DORIS)	<ul style="list-style-type: none"> - fréquence (rupture) - dépendances
13	Centralité ponctuelle de thématiques anciennes	Apparition de plusieurs termes thématiquement liés dans le corpus, qui ne sont pas des néologismes	Thématique de l'infrarouge (TTVS) Thématique des miroirs (TTVS)	<ul style="list-style-type: none"> - fréquence (formes chrono-homogènes) - variantes - dépendances (séries paradigmatiques)
14	Retrait de thématiques (anciennes mais non obsolètes)	Disparition de plusieurs termes thématiquement liés dans le corpus, qui ne	Thématique de l'étalonnage (TTVS)	<ul style="list-style-type: none"> - fréquence (formes chrono-

		sont pas obsolètes		homogènes) - dépendances (séries paradigmatiques)
15	Modification des connaissances à propos d'un concept donné	Changement de point de vue	<i>parasites</i> (TTVS) <i>opérateur</i> (DORIS)	- dépendances - fréquence
16	Modification de la structure des documents	Apparition de termes dans le corpus, mais qui ne sont pas des néologismes	<i>terminal de transfert</i> (DORIS)	- fréquence (rupture)
17	Mise à jour de la documentation qui constitue le corpus	Apparition/disparition de termes, de variantes, de thématiques	Exemple de la thématique des colles (TTVS)	- fréquence